AMERICAN JOURNAL of PHYSICS

A Journal Devoted to the Instructional and Cultural Aspects of Physical Science

VOLUME 17, NUMBER 4 APRIL, 1949

The Dimensions of Physical Concepts

PARRY MOON
Massachusetts Institute of Technology, Cambridge, Massachusetts

AND

DOMINA EBERLE SPENCER
Brown University, Providence, Rhode Island

DESPITE the extensive literature on the dimensions of physical quantities, the subject is still in a controversial state. In the earlier decades of this century, the trouble was associated with vestiges of the nineteenth-century belief that all physics could be reduced to mechanics and consequently all dimensional formulas could be expressed in terms of l, m, t. Today, physicists generally accept the idea that dimensions are arbitrary,[1] yet traces of the old belief still appear.

Maxwell[2] seems to be largely responsible for the emphasis on l, m, t as fundamental dimensions. Also, since he was interested in dimensions primarily as a means of changing from one system of units to another, he stressed *units* rather than *concepts*. His treatment has been followed by writers on the subject for the past 80 years, despite the fact that further developments have made change of units one of the least important of dimensional applications.

In the modern treatment of dimension theory, the primary application is the designation of concepts and the emphasis throughout is on concepts rather than units. Yet the older idea—that dimensions are associated with units—is often expressed and thus it seems worth while to discuss this basic question.

The viewpoint that we wish to examine is most precisely stated by Birge,[3] and we shall therefore refer to his papers in some detail. Before considering his arguments, however, we shall attempt to express our ideas on the subject and will then compare them with Birge's.

I. Concepts

In the first place, we agree with Bridgman[1] and Birge[3] on the relativity of dimensions—the *number and choice of primary dimensions are arbitrary*. As a corollary to this theorem, one may say that there is nothing "wrong" with the l, m, t system, the use of fractional exponents, or other controversial proposals. Dimension theory is a man-made device whose justification is its utility. The question therefore is not the rightness or wrongness of a dimensional idea but rather the *usefulness* of the idea. We shall employ

[1] P. W. Bridgman, *Dimensional analysis* (Yale Univ. Press, New Haven, 1931).

[2] See J. C. Maxwell and F. Jenkin, "On the elementary relations between electrical measurements," British Assn. Report for 1863, p. 130; Reports of the Committee on Electrical Standards, London, 1873, p. 59; J. C. Maxwell, *A treatise on electricity and magnetism* (Oxford Univ. Press), Vol. I, p. 1.

[3] R. T. Birge, "On the establishment of fundamental and derived units with special reference to electric units," *Am. J. Physics* **3**, 102, 171 (1935); "On electric and magnetic units and dimensions," *Am. J. Physics* **2**, 41 (1934).

this pragmatic criterion throughout and shall not attempt any metaphysical speculation.

The existence of physical *concepts* is evidently accepted by all physicists, even though they may deny that concepts have dimensions. The operational definition of concepts is also widely accepted. Bridgman[4] says,

> . . . we mean by any concept nothing more than a set of operations; the *concept is synonymous with the corresponding set of operations.*

Again he says,

> There probably is no statement either in Einstein or other writers that the [operational] use of 'concept' has been self-consciously made, but that such is the case is proved, I believe, by an examination of the way concepts are now handled by Einstein and others. For, of course, the true meaning of a term is to be found by observing what a man does with it, not by what he says about it. We may show that this is the actual sense in which concept is coming to be used by examining in particular Einstein's treatment of simultaneity.

The concepts of physics may be classified as *primary* and *secondary*. In mechanics, the primary concepts are usually taken as length, mass, and time; and an attempt is made to express everything in terms of these three primary concepts. This choice, however, is arbitrary and, even for mechanics, is not always most convenient. In heat and in electromagnetism, there is a distinct advantage in introducing additional primary concepts.

II. Designation Symbols

In classifying and correlating the various concepts of physics, one finds it helpful to introduce certain *designation symbols*, usually called dimension symbols. The thesis of this paper is that *these designation symbols apply to the concepts themselves and not to the units*. The symbols are primarily useful in showing at a glance the relations among concepts, irrespective of the choice of units and standards.

For instance, if l, m, t are chosen as primary concepts, then the concept of velocity is designated by the symbols lt^{-1}, force is lmt^{-2}, and power is l^2mt^{-3}. Or one may write the designation symbols in the form of an ordered set[5] or number-triple: for example, $[1, 0, -1]$, $[1, 1, -2]$, $[2, 1, -3]$. In all cases, the symbols designate the concepts, not the units.

In treating any branch of physics, one writes down a list of the concepts and attaches to each a designation symbol or group of symbols. This fixes the particular concepts for the branch of physics and shows how these concepts are related. Up to this point, no consideration has been given to units. After this aspect of the subject is settled, however, one is free to proceed to the independent consideration of (i) names of concepts, (ii) choice of units, (iii) names of units, (iv) choice of standards.

The foregoing procedure was used in revising the concepts of radiometry and photometry.[6] As stated, the primary object of introducing designation symbols is to aid in the classification of concepts. It is found, however, that the symbols have value in other applications such as the checking of equations, the derivation of new equations, and the solution of problems by model theory.

III. Principles

Consider the establishment of a satisfactory set of designation symbols for physics. In view of the controversial nature of the subject, it seems advisable to state as precisely as possible the postulates on which one may proceed:

(1) The choice of primary concepts is arbitrary. The number of primary concepts, as well as their nature, may be varied at will. The designation symbols will change with the choice of primary concepts.

(2) The concepts that are chosen as primary have no necessary connection with the quantities that are fixed by standards. For instance, the fact that a standard ohm is a convenient piece of laboratory equipment does not mean that *charge* or *permittivity* may not be taken as primary concepts, with resistance as a derived quantity.

(3) Designation symbols for concepts can be

[4] P. W. Bridgman, *The logic of modern physics* (Macmillan Co. New York, 1927).

[5] See, for instance, L. E. Dickson, *Linear algebras* (Cambridge Univ. Press, 1930).

[6] Parry Moon, "A system of photometric concepts," *J. Opt. Soc. Am.* **32**, 348 (1942). P. Moon and D. E. Spencer, "A study of photometric nomenclature," *J. Opt. Soc. Am.* **36**, 666 (1946); "Internationality in the names of scientific concepts," *Am. J. Physics* **14**, 285 (1946); **14**, 431 (1946); **15**, 84 (1947); "Utilizing the mks system," *Am. J. Physics* **16**, 25 (1948); "Modern terminology for physics," *Am. J. Physics* **16**, 100 (1948).

fixed without regard to the specific units that are employed.

(4) With a given choice of primary concepts, a 1:1 correspondence should exist between concepts and their designation symbols.

(5) The number of primary concepts should be a minimum, consistent with the satisfaction of postulate (4).

The principal requirement is the 1:1 correspondence between concepts and designation symbols, stated in postulate (4). In the l, m, t system, this principle is violated in a number of cases. For instance, on the electrostatic basis *capacitance* has the same designation symbol as *length*, and *resistivity* has the same symbol as *time*. Everyone is familiar also with the identical designation symbols for *energy* and *torque*. Yet capacitance and length are universally regarded as distinct concepts. The same is true of resistivity and time, energy and torque. The ambiguity in their designation tends to introduce into dimension theory a vagueness that is anything but helpful.

A basic principle of dimension theory, stated by FOURIER in 1822 and by MAXWELL in 1863, is the principle of *dimensional homogeneity:* if two terms in an equation of physics have the same dimensions, they can be added; if not, they cannot be added. But no one seems to have remarked that this law cannot possibly hold universally unless there is a 1:1 correspondence between designation symbol and concept. In the l, m, t system, the principle holds sometimes but not always; and it is these exceptions that make traditional dimension theory so treacherous and unsatisfactory a tool.

By taking temperature and electric charge as additional fundamental concepts, we remove much of the ambiguity in the subjects of heat and electromagnetism. If now two fundamental space concepts,[7] designated by l_r and l_t, are introduced in place of the usual length l, all discrepancies are eliminated and a true 1:1 correspondence is obtained between concepts and designation symbols. This subject will be treated in detail in another paper.[8]

[7] The idea of splitting the usual l-dimension into three was proposed by W. Williams in 1892 (Phil. Mag. **34**, p. 234). Our use of *two* primary lengths appears to be new.

[8] P. Moon and D. E. Spencer, "A modern approach to dimensions," J. Franklin Inst. (1949).

IV. Geometric Representation

The equations of physics are mathematical equations giving relations among numbers. Physically they are meaningless unless supplemented by a text that tells what the symbols signify and under what circumstances the equations apply.

Sometimes it is convenient to include part of the text in the equation, and one way of accomplishing this is by writing equations in terms of complete physical quantities instead of pure numbers. A *physical quantity* or *physical magnitude* is a specific amount of a definite kind, as a *potential difference of 120*. We shall denote such a quantity by a symbol in curly brackets,[9] as $\{A\}$, and shall understand this symbol to designate both the magnitude (as 120) and the concept (as potential difference). Then

$$\{A\}=A[a^1, a^2, \cdots a^k]. \tag{1}$$

Here the magnitude of $\{A\}$ is A, while the kind of physical quantity is specified by the designation symbol, which may be written[10]

$$[a^1, a^2, \cdots a^k]=a^i, \quad i=1, 2, \cdots k.$$

The quantity $\{A\}$ may be called an *idon* (from Gk *εἶδος* meaning class, species),[11] and the study

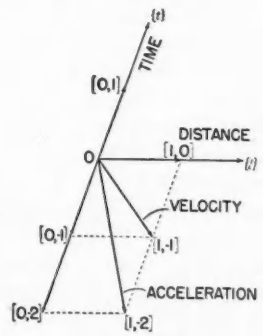


FIG. 1. An idonic 2-space, showing the time and distance axes, as well as base vectors for time [0, 1], distance [1, 0], velocity [1, −1], and acceleration [1, −2].

[9] P. Moon and D. E. Spencer, "A geometric treatment of dimensions," to be published.

[10] J. A. Schouten and D. J. Struik, *Einführung in die neueren Methoden der Differentialgeometrie* (P. Noordhoff, Groningen, 1935), Vol. I.

[11] A long and unsuccessful attempt was made by the authors to employ the traditional words *dimensions* and *dimensional analysis*. But ambiguities constantly arose and required all sorts of clumsy circumlocutions before any semblance of clarity and precision could be obtained.

of these quantities may be termed *idon theory* or *idon analysis*.

Any equation of physics, such as

$$F=ma, \tag{2}$$

can be written in extended form by merely substituting idons for the usual numerical magnitudes. The foregoing equation then becomes (for l, m, t as primary concepts),

$$\{F\}=\{m\}\{a\}.$$

If $\{F\}=Fa^i$, $\{m\}=mb^i=m[0, 1, 0]$, $\{a\}=ac^i=a[1, 0, -2]$, then

$$Fa^i=(ma)\cdot[1, 1, -2], \tag{3}$$

where the product of two designation symbols is obtained by adding corresponding coordinates (Sec. V). Equation (3) may now be split into two parts, if desired. The scalar part is Eq. (2), while the idon vector part gives the designation symbol for force:

$$a^i=[1, 1, -2].$$

Geometrically, an idon $\{A\}$ may be considered as a *vector* in a k-dimensional manifold. The magnitude of the vector is denoted by A and its direction is specified by the *base vector* a^i. The numbers a^1, $a^2\cdots a^k$ are the *coordinates* of the base vector. The fictitious space in which the idon vectors lie is called *idon space*. Its dimensionality k is equal to the number of primary concepts employed in the problem under consideration. Thus if l, m, t are used as primary concepts, idon space will be a 3-space and the coordinate axes will be l, m, and t.

For kinematics, two primary concepts, l and t, are ordinarily employed. In this case, idon space is 2-dimensional (Fig. 1). Any physical distance is represented by a vector $\{l\}$ along the l-axis:

$$\{l\}=l[1, 0],$$

and any value of time is represented by a vector along the t-axis:

$$\{t\}=t[0, 1].$$

A specific velocity is an idon vector[12] in another direction (Fig. 1):

$$\{v\}=v[1, -1],$$

while a definite acceleration is

$$\{a\}=a[1, -2].$$

Evidently each species of physical quantity is associated with a definite *direction* in idon space. This direction is specified by the designation symbol $[a^1, a^2, \cdots a^k]$ which represents a base vector a^i with one end fixed at the origin. Changing the magnitude A of a physical quantity merely stretches the $\{A\}$-vector in the direction determined by the base vector a^i.

Since the length of the vector $\{A\}$ is always given by A, irrespective of what coordinates the base vector may have, a^i evidently acts as a *unit vector*. Note, however, that unit vectors in different directions cannot be compared. For instance, the base vectors $[1, -1]$ and $[1, -2]$ of Fig. 1 do not *look* like unit vectors, but they *are* unit vectors. The difficulty disappears when one realizes that idon space is not euclidean but is an *affine space*. Idon space has no metric, and angles in this space have no absolute significance.

V. Algebra

The algebra of idons may be summarized as follows:

(1) *Addition*. Given two idons,

$$\{A\}=Aa^i \quad \text{and} \quad \{B\}=Bb^i.$$

Then addition is defined only if $\{A\}$ and $\{B\}$ are collinear.[13] If this condition is satisfied,

$$\{A\}+\{B\}=(A+B)a^i=(A+B)b^i. \tag{4}$$

Subtraction is included as a special case of Eq. (4). Addition is commutative and associative.

(2) *Multiplication by a scalar*. Given an idon $\{A\}$ and a scalar k, then the product is

$$k\{A\}=kAa^i. \tag{5}$$

Particularly annoying were references to the various *dimensions* (lengths in different directions) of models and their relation to the *dimensions* of the prototype device, also references to the dimensionality of dimension-space (with l, m, t as fundamental dimensions, we have a 3-*dimensional* dimension-space). Of course, the word dimensions could be enclosed in quotation marks whenever it occurs with a special meaning. This has not been done in the present paper. The only satisfactory solution seems to be the introduction of a new word in place of dimensions. For this word, we select *idon*.

[12] One must, of course, distinguish between a vector in idon space and a vector in the euclidean space in which the physical phenomena take place. If the physical quantities are vectors or tensors, we consider only their magnitudes. This procedure introduces no limitations on the generality of idon theory.

[13] The principle of dimensional homogeneity.

Multiplication by a scalar is commutative: $k\{A\}=\{A\}k$.

(3) *Multiplication.* The product of two idons is defined as

$$\{C\}=\{A\}\{B\}=AB[a^i+b^i], \tag{6}$$

where the base vector of $\{C\}$ is obtained by adding the coordinates of the base vectors for $\{A\}$ and $\{B\}$. Multiplication is commutative and associative, and is distributive with respect to addition.

(4) *Division.* The quotient of two idons is

$$\{D\}=\frac{\{A\}}{\{B\}}=\frac{A}{B}[a^i-b^i]. \tag{7}$$

These four operations are the only ones in idon theory. They are called *allowable operations.* An idon can be raised to an integer power by repeated use of Eq. (6), and a natural extension to non-integer powers results in the equation,

$$\{A\}^\gamma=A^\gamma[\gamma a^i]. \tag{8}$$

Note that the operations are similar (though not identical) with the familiar operations on complex numbers in polar form.

VI. The Π Theorem

Since there are only four allowable operations in idon theory, the most general idonic equation is of the form,

$$\sum_{i=1}^{j} K_i\{A\}^{\alpha i}\{B\}^{\beta i}\cdots\{N\}^{\nu i}=0, \tag{9}$$

where K_i are scalar coefficients, which may be transcendental functions of idonic scalars. Examination of the equations of physics shows that all of them can be written in this form. If transcendental functions occur in physical equations, the arguments of these functions can always be written as idonic scalars (dimensionless quantities); and if the argument of a function is an idonic vector (quantity having dimensions) the function can be at most only a power function. Thus *Eq. (9) represents the most general equation of physics.*

Sometimes it is convenient to express the most general equation of physics in an alternative form. First Eq. (9) is divided by one of the terms, to obtain an expression of the form,

$$\sum k_l\{A\}^{\rho l}\{B\}^{\sigma l}\cdots\{N\}^{\omega l}+1=0. \tag{10}$$

Since addition is allowable only for quantities having the same designation symbols, and since 1 is an idonic scalar, all other terms must likewise be idonic scalars (dimensionless quantities). Therefore any equation of physics may be expressed in terms of idonic scalars having the form,

$$\Pi_i=\{A\}^{\rho i}\{B\}^{\sigma i}\cdots\{N\}^{\omega i} \tag{11}$$

where $i=1, 2, \cdots k$. The equation may contain any functions of these products, so the most general equation of physics may be written,

$$\varphi(\Pi_1, \Pi_2, \cdots \Pi_m)=0. \tag{12}$$

This is the Π *theorem*, first stated by Vaschy[14] in 1896 and later by Riabouchinsky (1911).

VII. Utility of Idon Theory

Idon theory is useful in six definite ways:

(a) Designation of physical concepts. Each concept is specified by a designation symbol $[a^1, a^2, \cdots a^k]$ which may be represented geometrically as a base vector a^i in a k-dimensional affine space.

(b) Derivation of equations in terms of idonic scalars Π, for physical problems that are so involved that detailed analysis is impracticable (see reference 9).

(c) Model theory based on the use of idonic scalars Π.

(d) Change from one system of units to another. In idon theory, this change is effected by a transformation of the A's, the base vectors being unchanged.

(e) Transformation of axes in idon space corresponding to a change in fundamental concepts. This change is effected by an ordinary affine transformation of the base vectors,

$$a^{i'}=A_i{}^{i'}a^i. \tag{13}$$

(f) Checking of equations. Each term of an equation of physics must have the same base

[14] A. Vaschy, *Théorie de l'électricité* (Paris, 1896), p. 13; D. Riabouchinsky, "Méthode des variables de dimension zéro et son application en aérodynamique," *L'Aérophile* 19, 407 (1911).

vector.[15] This is a necessary but not a sufficient condition for the validity of the equation.

These six applications are all handled by idon theory as straightforward results of the methods of affine geometry. The primary emphasis is on (a) the *designation of physical concepts*, from which the other applications follow as a matter of course. The only place where *units* enter is in (d), which is the rather trivial case of a transformation of scales along the coordinate axes. It would seem that the association of dimensions with units rather than with concepts is an old-fashioned idea, promulgated by those whose chief use of dimensions was in changing from one system of units to another. The other five applications, however, are more important than (d) and they can be expressed most simply and precisely on the postulate that designation symbols (dimensions) are associated with physical *concepts*.

VIII. Comparison

We now compare the idonic idea with some statements on dimensions from the papers of Birge.[3] On p. 102 of the *American Journal of Physics*, vol. 3 (1935), he says,

> Physical **units** are divided into two classes, *fundamental* or *primary* **units**, and *derived* or *secondary* **units**. Most of the difficulties experienced in this subject are connected with the various possible ways that have been used, or may be used, to obtain derived **units**. I shall start, however, with a brief discussion of fundamental **units**.
>
> A fundamental **unit** is distinguished by the fact that it is entirely independent of all other **units**, in respect to both its *magnitude* and to its *dimension* or *dimensions*. It is customary to assign to each fundamental **unit** a specific dimension. Thus the **unit** of length is said to have the dimension of length. Because, however, of the arbitrary character of dimensions, as presented so ably by Bridgman, the choice and number of fundamental **units** are arbitrary.

The word "unit" has been printed in bold-face type every time it occurs, to indicate that we would replace it by *concept* in each case. With such a change, the Birge quotation becomes an admirable statement of the views expressed in the present paper. On p. 103, however, Birge employs the word "unit" with respect to the establishment of cgs standards. Thus, according to our viewpoint, Birge's use of the word is somewhat ambiguous. Of course, either way is permissible, the question being merely one of utility.

On p. 105, Birge says,

> Since angle is the ratio of two lengths, its *unit* is necessarily dimensionless. Thus it is possible to have more than one dimensionless unit for certain quantities by the simple artifice of using simultaneously units of different magnitude for the same sort of primary quantity (in this case *length*) in the defining equation.

We would say that designation symbols (dimensions) are not associated with the unit at all but with the concept. The physical quantity, *angle*, may be written

$$\{\theta\} = \theta[-1, +1]$$

where the first coordinate of the base vector refers[8] to radial length l_r and the second refers to tangential length l_t. In this way, angle is a vector in idon space, not a scalar, and the designation symbols for energy and torque are distinct. Thus we differ from Birge's statement in three particulars: (a) Dimensions are not associated with units, (b) the concept of angle[8] does have dimensions, (c) the use of different units for l_r and l_t has nothing to do with dimensions.

On p. 105, Birge considers the concept of density, as defined by the equation,

$$d = m/V:$$

> The dimensions of d are necessarily those of the unit, and it is to the *unit*, not to the physical quantity, that one assigns a dimension or dimensions . . . one chooses primary and derived units but not primary and derived quantities.

On the contrary, we write the complete equation corresponding to the above as

$$\{d\} = \{m\}/\{V\} = m/V[-3, 1, 0]$$

if l, m, t are taken as primary concepts. Obviously, this equation holds as a definition of the *concept of density*, irrespective of the units that may be employed.

In common with some philosophers, Birge[3] says (p. 105), "Every equation used in science . . . gives a relation between *numerical measures*

[15] It is, of course, possible to manufacture artificial equations that are not idonically homogeneous. But every equation of physics can be written in homogeneous form, provided that a 1:1 correspondence exists between designation symbols and the concepts used in the equation.

of certain quantities, not between the physical quantities themselves." As indicated previously, this statement is true in a sense, but it gives a very superficial view of the subject. Bridgman points out in *The Logic of Modern Physics*[4] that the equation is nothing when divorced from its explanatory text. We have shown how some of the text can be incorporated into the complete equation by representing physical quantities by idons.

We conclude with a few quotations indicating the failure of the usual dimension theory as a method of designating concepts. On p. 106, Birge[3] says, "Bridgman's theory of dimensions stresses the fact that the dimensions of a unit give us no information about the intrinsic nature of the physical quantity of which it is a unit."

From p. 179, we quote:

> . . . the intrinsic character of a physical quantity has no necessary connection with the *dimensions* assigned to the *unit* of that quantity. This is the central theme of the present paper, and of the work of Bridgman, Planck, H. Abraham, and other writers. The I.E.C. seems to favor the theory of the absolute character of dimensions. Having voted that B and H were of different physical character, the I.E.C. apparently assumed, without further debate, that they *therefore necessarily* had different dimensions.

From the same page,

> If the entire scientific world could only realize clearly that the character of a physical quantity is a matter of philosophy, while the assigned dimensions of a unit are a matter of mere convention having only arithmetic significance, unending and utterly fruitless controversies would be avoided.

As we understand these quotations, they can be summarized by saying that there is no necessary connection between designation symbols (dimensions) and physical quantities. Since the choice of primary concepts is arbitrary, this statement is a truism. But a fundamental aim of the method should be to so choose the designation symbols that, for a given set of primary concepts, *there is a 1:1 correspondence between concept and symbol.* When this desideratum is realized, the source of complaints, that "dimensions give no information about the nature of the physical quantity," is eliminated.

IX. Conclusions

The thesis of this paper is that *dimensions are associated with concepts, not with units.* Bouasse[16] says, "the collection of stupidities formulated in connection with dimensions truly exceeds the limit of what is reasonably allowable." The association of dimensions with units is not a stupidity.[17] It is not incorrect, but also it is not as useful as the modern treatment.

There are six principal applications of idon theory (dimensions), all of which are handled effectively and surely by the modern treatment and somewhat precariously by the older methods. In particular, the change of units, which is emphasized in the usual treatments, becomes merely a trivial case of coordinate transformation by the new method.

The modern approach utilizes an extension of the usual equation of physics. A physical quantity is represented by a vector in a k-dimensional affine space. The magnitude of the quantity is represented by the length of the vector, while the species of physical quantity is specified by a base vector whose coordinates determine the direction of the vector in idon space. Such a treatment tends to purge the subject of much of its former vagueness and gives a geometric picture that aids in the visualization of idonic problems.

In conclusion, we sincerely thank those who have read and commented on previous drafts of this paper: R. T. Birge, P. W. Bridgman, Royal M. Frye, J. C. Hunsaker, Phillippe Le Corbeiller, V. F. Lenzen, Duane Roller, J. A. Stratton, and Dirk J. Struik.

[16] H. Bouasse, *Cours de magnétisme et d'électricité* (Paris, 1914), Vol. I, p. 420.

[17] Other examples of the conventional treatment are H. Abraham, "A propos des unités magnétiques," Nat. Res. Council, *Bulletin*, **93**, 8 (1933); J. Wallot, "Zur Theorie der Dimensionen," *Zs. f. Physik* **10**, 329 (1922); V. F. Lenzen, "The meaning of dimensions," *Am. J. Physics* **9**, 245 (1941); J. C. Hunsaker, "Dimensional analysis and similitude in mechanics" (in *Theodore von Kármán anniversary volume*, Pasadena, 1941).

Lecture Demonstration of Longitudinal Waves

W. W. Sleator
University of Michigan, Ann Arbor, Michigan

A FIRST step in one method of deriving the equation

$$U=(E/\rho)^{\frac{1}{2}} \tag{1}$$

for the velocity of plane longitudinal waves is a proof of the equation

$$u/U=-\Delta V/V, \tag{2}$$

where u is the velocity of any layer of the material, U is the velocity of the waves, and $\Delta V/V$ is the change in volume per unit of original volume, or the strain, at that layer. This relation holds throughout the wave train. If the wave velocity U can be qualified as positive when, for example, it is directed away from an origin and toward the right, and negative when directed toward the left, then Eq. (2) holds for waves traveling in either direction. The apparatus to be described makes possible a very easy proof of Eq. (2).

Figure 1 shows 30 wood disks some 4 in. in diameter and $\frac{5}{16}$ in. thick, that slide easily on a wooden rod about 3 feet long. These and other numerical dimensions are not important. There are two small screws at the ends of a diameter in the edge of each disk, and similar wire links, made out of paper clips, tie each disk to its two neighbors, so that they can be pushed into contact, forming a solid column, or pulled apart so as to form an open row with the disks equally spaced and about $\frac{3}{8}$ in. apart. The demonstrator starts with the open row (Fig. 1(a)), and takes the left end disk in one hand, sliding it uniformly along the rod so as to gather disk after disk into a solid column (Fig. 1(b)). As this goes on he points with the other hand, as indicated, to the front end of the solid column, where new contacts are taking place, and emphasizes the two different velocities involved. The pushing hand and the moving material have the velocity u; the end of the solid column, which is seen to be advancing through the row as disk after disk is collected, has the greater velocity U. Here u and U are in the same direction.

If the thickness of any disk is denoted by a and the free space between is b, then it is obvious that as the actual wood advances a distance b the head of the solid column collects one more disk and moves forward the distance $a+b$, so that

$$u/U=b/(a+b). \tag{3}$$

Now, b is the space closed up as the column solidifies and $a+b$ is the corresponding space originally occupied. Hence

$$b/(a+b) = (\text{change in volume})/(\text{original volume}). \tag{4}$$

Here a and b are taken as positive numbers, and there results, from Eqs. (3) and (4),

$$u/U=-\Delta V/V,$$

a result identical with Eq. (2).

One advantage in having the disks connected by similar links is that, with the disks all in contact, it is easy to distribute them into a uniform open row, as they should be when the demonstration begins. However, a second and more important advantage is that with these connections the experiment can be done in reverse, so as to illustrate the progress of a rarefaction. As in Fig. 1(c), the demonstration is started with a solid column. Taking the left end disk in one hand the operator draws it away from the rest, and a rarefaction is seen to advance into the compact row (Fig. 1(d)). He points to the end of the open column, which has the mean velocity U, while the moving wood and the pulling hand have the velocity u, pointing out that u and U are now opposite, U toward the right and u toward the left. Here while the wood moves a distance b, the surface separating compact from open formations moves only the distance a. But as before,

$$b/a = (\text{change in volume})/(\text{original volume}). \tag{5}$$

We take a and b as positive numbers and the change in volume is an increase, so that ΔV is positive, but since u is now opposite to U we have, as before,

$$u/U=-\Delta V/V.$$

Hence, Eq. (2) holds as a condensation advances into an open column, and also as a rarefaction advances into a compact column.

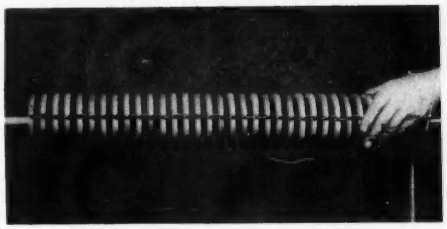
Half of this demonstration can be performed with equipment still simpler. One distributes similar two-inch washers uniformly along a horizontal rod. Then, grasping the end washer he pushes it along and gathers the others one after another into a solid moving column. This demonstrates the advance of a condensation, and the quantitative relation is obvious, but time is required in order to redistribute the washers, and one cannot reverse, so that the connected disks are decidedly better. They work both ways.

A student inclined to complain might raise an objection to the disk experiment, saying that it showed only one velocity u for the moving material, instead of a different velocity for each layer, as in an actual wave, where, moreover, the velocity of each layer is variable. It can be replied that the disk outfit gives a magnified slow-motion picture of what happens momentarily in an actual harmonic wave train. For we have taken any minute displacement of a piston generating real longitudinal waves, and continued it at constant velocity in time and space, so that our slow perceptions can observe the essentials of the process. The strain $\Delta V/V$ is not the same everywhere in harmonic waves, as it is in the solid column of disks, but the actual strain anywhere along the real wave train is related to velocity where that particular strain is, as shown by Eq. (2). What, indeed, may one ask of a lecture experiment if not that it should clear away nonessentials and reveal important things, such as Eq. (2), in simplicity?

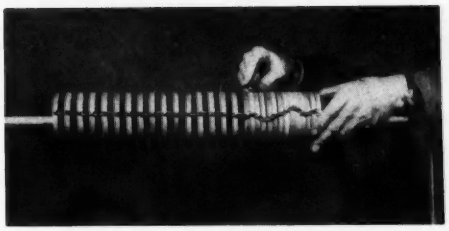
If we preserve the left end of the open row as origin of the x coordinate, but begin the collecting process at the right end we can think of U as a negative velocity, and then, when u is negative also a condensation advances into the open row and ΔV is negative, so that, as before,

$$u/U = -\Delta V/V.$$

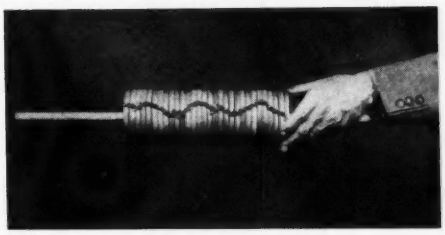
Similarly, when a rarefaction moves through the compact column toward the origin, u is positive and U negative but the change in volume is an increase so that the same relation holds, and $u/U = -\Delta V/V$ in any case. The general principle is that u and U are in the same direction in a condensation, opposite in a rarefaction. This fact, naturally, is independent of the direction arbitrarily adopted as positive along a coordinate axis.



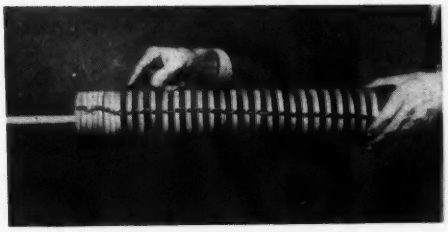
(a) Starting



(b) The Compression



(c) Starting



(d) The Rarefaction

FIG. 1. Apparatus for demonstrating longitudinal waves, showing: (a) disks uniformly spaced at start of compression; (b) during the compression two velocities u and U are apparent; (c) disks in contact at the start of rarefaction; (d) during the rarefaction velocities u and U are in opposite directions.

There are several ways to complete the proof of Eq. (1) after Eq. (2) is established[1] but perhaps the simplest is to use Newton's second law. While a condensation is advancing into the open column (of cross section A) the mass of the moving material is increasing at the rate $A\rho U$, in grams per second, for example, and the momentum is increasing at the rate R where

$$R=A\rho Uu. \tag{6}$$

However, the force required to produce this rate of increase is $A\Delta P$, where ΔP is the excess pressure in this condensation, over and above the normal pressure in the undisturbed material. Hence, by the second law of motion

$$A\Delta P=A\rho Uu=A\rho U^2u/U, \tag{7}$$

or

$$\Delta P=\rho U^2u/U. \tag{8}$$

But we have just seen that u/U is $-s$, where s is the strain corresponding to the stress ΔP. If the coefficient of elasticity appropriate to the process involved is the positive constant E,

$$E=-\Delta P/s=\rho U^2. \tag{9}$$

Hence

$$U=(E/\rho)^{\frac{1}{2}},$$

which was to be proved.

We can readily describe the motion of the disks in the general manner appropriate to any kind of waves. To begin, let us agree that by saying that a quantity represented by a letter is positive, we mean that a positive number must be substituted for that letter in numerical calculation, and a negative number for it if it is negative. We take the positive x coordinate as the axis of the disks and the origin at the initial position of the left end disk. At the time $t=0$ we begin to move that disk toward the right with the constant velocity u. Then at any time t the displacement of the first disk is y_0 and

$$y_0=ut. \tag{10}$$

At any x farther along the row the displacement at the same time t is y_x and

$$y_x=u(t-x/U). \tag{11}$$

Here we are supposing that t is greater than x/U so that the disk at x is among those already picked up and in motion at the time t. This means that the disk at x began to move after the first was started, that is x/U later. Here U, being toward the right, is positive. Then

$$\partial y_x/\partial x=-u/U. \tag{12}$$

In case u is positive there is a condensation in the row because, of two disks, the one nearer the origin began to move first, and has moved farther than the other, so that any two disks are closer together than they were in the original open column. In other words

$$\partial y_x/\partial x=\text{strain}=\Delta V/V, \tag{13}$$

and since $\Delta V/V=-u/U$, and both u and U are positive, ΔV is negative and the strain is a condensation. The case where u is negative, treated in the same way, yields a rarefaction in the column.

If U is negative, the change is moving toward the origin. Let the disk at $x=0$ begin to move at $t=0$ so that

$$y_0=ut. \tag{14}$$

Then at the same time t the displacement of a disk originally at x is y_x, and y_x is what the displacement at $x=0$ will be at a time t' later than t, so that, at x and t,

$$y_x=u(t+t'). \tag{15}$$

Here t' is a positive interval or increment of time, and x is positive, so that $t'=-x/U$. Therefore

$$y_x=u(t-x/U). \tag{16}$$

Hence as before

$$\partial y_x/\partial x=-u/U. \tag{17}$$

All special cases conform to the general principle: in a condensation u and U are in the same direction, in a rarefaction they are opposite.

It is obvious that, using partial derivatives, $u=\partial y/\partial t$, and Eq. (2) may be written

$$U=-(\partial y/\partial t)/(\partial y/\partial x). \tag{18}$$

Lindsay[2] has pointed out that this equation is not a general differential equation of wave motion, for, he says, "It holds only for a wave in the positive x direction." It is evident that the

[1] W. W. Sleator, "Proofs of the equation $U=(E/\rho)^{\frac{1}{2}}$ for the velocity of sound," *Am. J. Physics* **17**, 51 (1949).

[2] R. B. Lindsay, *General physics* (Wiley, 1940), p. 388.

distinction indicated by the plus and minus signs is not adequate to distinguish one from another the innumerable different directions in which waves can be propagated in space, and when one deals with waves in three dimensions, it is convenient to regard the wave velocity U as a universal positive constant. In the same way, in our evaluation of the trigonometric functions of angles terminating in different quadrants, we regard the radius, or hypotenuse of the triangle, as universally positive. But while we are dealing with plane waves and are at liberty to choose an x axis as the axis of propagation, normal to the wave fronts, Eq. (18) shows no lack of generality. The reversal of both velocities does not change the sign of the strain.

What About Friction?

FREDERIC PALMER
The Franklin Institute Laboratories for Research and Development, Philadelphia, Pennsylvania

THIS article presents a brief resumé of the present status of sliding friction obtained from a digest of more than 350 references to the literature.* The most important phenomena only are treated, some of which form the basis of experimental work now in progress at the laboratories of the Franklin Institute, Philadelphia. The subject matter is divided as follows:

Part I

The classical laws of friction are stated and various quotations from the works of Leonardo da Vinci, Amontons, Coulomb, Euler, Morin, and Bikerman are given to show how the laws were derived and to what extent they are valid.

Part II

The molecular theory of friction, as expounded by Bowden and his associates, is described and discussed. Photographic evidence is presented of the existence of the area of actual contact, the welding of copper to steel during dry sliding, and the plowing of the softer surface by the harder.

The dependence of friction on velocity is discussed and applied to the phenomenon of stick-slip.

Part III

The electrical theory of friction is discussed and an attempt is made to explain the action of the high speed friction saw. Some photographic evidence is presented.

The dependence of friction upon the presence of an adsorbed oxide layer is discussed together with a few facts about lubrication.

The present state of our knowledge of friction is briefly summarized.

* The author is indebted to Rita A. Erickson for collecting the references to the literature and summarizing many of the papers.

PART I

Classical Laws

1. *Introduction*

Every student when studying college physics for the first time comes to regard the subject of friction with mixed feelings. He finds that the motion of one surface over another is always opposed by a force due to friction irrespective of the direction of movement, hence the efficiency of any mechanical device is reduced because of the friction of its moving parts. Corrections have to be applied in order to make calculated results agree with experience. This is a nuisance. Yet, without friction, he finds that nails, screws, and moving belts would be useless; trains and automobiles could not start, but if going, could not stop without a smashup; and his own ability to walk to his classroom would be lost. Thus, whether friction plays the part of devil or of angel, it exercises an influence upon everything that moves; it springs into action the instant a sliding force is applied; it may prevent motion that would otherwise occur (as when a body is at rest upon a plane inclined to the horizontal at a small angle).

The student learns further that (1) the force of static friction is greater than that of kinetic friction which, in turn, is greater than that of rolling friction; (2) the force of kinetic friction can be reduced by polishing and by lubricating the surfaces in contact; (3) the rubbing of one surface over another produces wear; (4) the

cause of friction is the intermeshing of microscopic surface roughnesses; (5) the "laws" of sliding friction (as customarily cited in textbooks) may be summarized as follows:

1. Frictional force is directly proportional to load, that is, to the total force which acts normal to the sliding surface.
2. Frictional force for a constant load is independent of the area of contact.
3. Frictional force is independent of the velocity of sliding.
4. Frictional force depends upon the nature of the materials in contact.

Further than this most students do not go; first, because of a lack of interest since friction is not an exciting topic full of fireworks, like atomic energy; second, because, even with the requisite interest, an inquiring student will be discouraged to find the answers to his questions scattered through the literature rather than brought together in easily available form.

What authority have we for the so-called "laws" of friction? What apparatus was used in carrying out the basic experiments? How accurate were the measurements? What deviations from these "laws" are now known? What alternative theories as to the nature of friction have been proposed? What theory is generally accepted today?

Such questions deserve straightforward, carefully prepared answers which the instructor in college physics cannot give without knowledge previously acquired as the result of a special interest of his own in friction. In what follows, it is proposed to supply at least partial answers to some of these questions, and to provide a short bibliography to be consulted by those who wish to carry their inquiry back to the original sources.

2. *The "Laws" of Friction*

a. Amontons.[1]—In presenting his paper entitled "De la Resistance Causée Dans les Machines" before the Royal Academy of Sciences in 1699, Guillaume Amontons said, in part:

"Indeed of all the authors who have written of the unstable forces, there is perhaps not one who has paid sufficient attention to the effect of friction in machines, and of the resistance caused by the stiffness of ropes, nor who has given us the rules for understanding the one and the other, and for reducing them to calculation; although indeed that understanding may not be as necessary for passing judgement on the action of a machine as is that of the different relations of the parts which compose it, yet, it is only too true that the lack of such an understanding of the resistance caused by friction and by the stiffness of the ropes is almost always in machines, a danger much more to be feared than its size, which, up to the present, has been unknown."

With the modesty characteristic of a true scientist, Amontons went on to describe the first experiments on the measurement of frictional resistance ever to be brought before an audience of professional men. His apparatus was simple and crude. A spring bore vertically down upon a block capable of sliding horizontally on a plane surface. A spring balance measured the horizontal force necessary to make the block move. The materials of the sliding surfaces were varied (iron, copper, lead, wood), as was their area.

Amontons concluded that the resistance caused by friction (1) is proportional to the force (load) with which the upper surface presses against the lower; (2) for the different materials used, is approximately $\frac{1}{3}$ of that force ($\mu = \frac{1}{3}$); and (3) is independent of the area of contact. He found that when the surfaces were greased the resistance is independent of the nature of the surfaces; and he believed that the resistance depends upon the velocity of movement. Thus the first two of the so-called "classical" laws of friction were stated for the first time publicly by Amontons in much the same terms as they are today. He proceeded to theorize as to the nature of friction as follows:

"Now if we consider carefully the nature of friction, we find that it is merely the action by which a body that is pressed against another is moved on the surface of that which it touches, and that as the surfaces which rub against each other cannot be considered either as rough and uneven or as perfectly smooth, it is not impossible in the first case that these unevennesses

[1] G. Amontons, "De la Resistance Causée Dans les Machines," *Mémoires de l'Académie Roy. Sci. de Paris* (1699), p. 206.

may have convex parts and concave parts, and that the former enter into the latter which produces a certain resistance when one wishes to make them move, since for lifting them there is needed as much force as that which presses one upon the other, and that the action of the unevennesses . . . is the same as that of the inclined plane used in raising a load, hence it follows that the greater the pressure, the more considerable is the resistance to movement."

Amontons does not distinguish between static friction and kinetic friction, and does not appear to see the significance of the angle of the inclined plane, mentioned above, as we now recognize it in the "angle of repose." He was, like the rest of the world, unaware that nearly two hundred years earlier the same basic principles had been recorded in the notebooks of Leonardo da Vinci.

b. Leonardo da Vinci.[2,3]—Da Vinci's notebooks contain, among others, the following items:

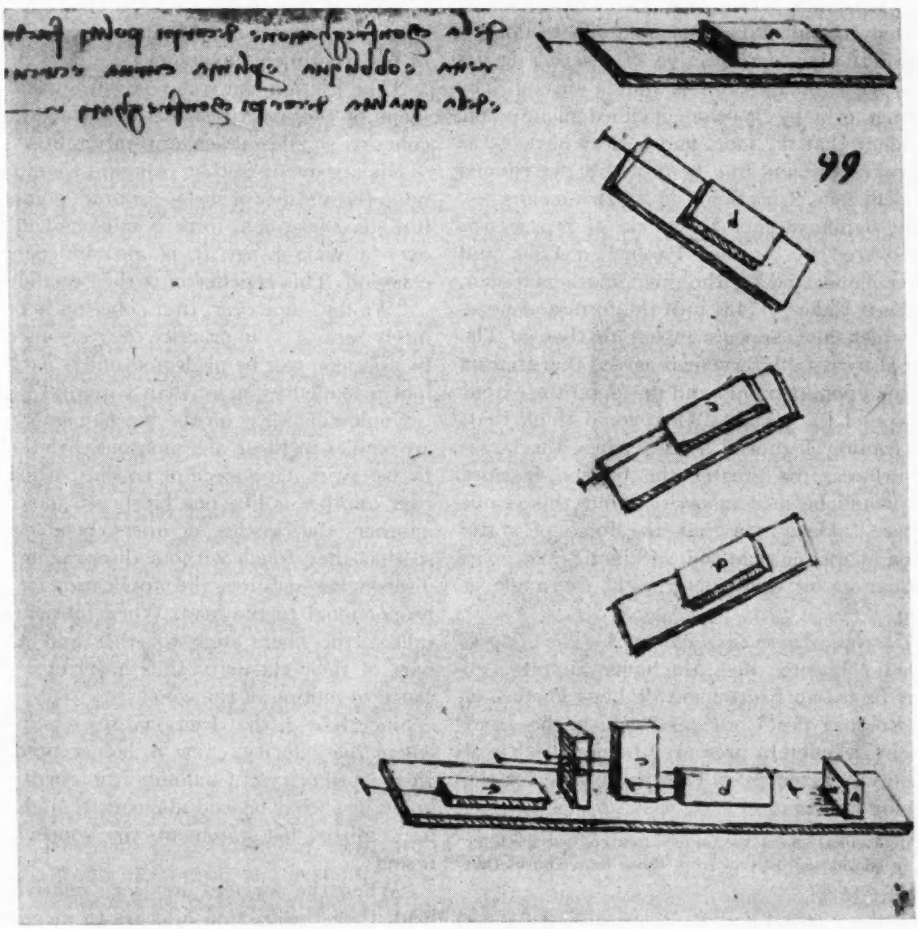


FIG. 1. Da Vinci's drawings illustrating his experiments on friction. Reproduced from the *Codice Atlantico* exactly as they appear in his notebook.

[2] L. da Vinci, *The notebooks of Leonardo da Vinci* (compiled by E. MacCurdy, Reynal and Hitchcock, 1938).
[3] L. da Vinci, *Il Codice Atlantico; Herausg. von der Accademia dei Lincei* (1508). Quoted by G. Vogelpohl, *Oel und Kohle* **36**, 89–93 and 129–143 (1940).

(1) Frictional resistance is proportional to roughness.

(2) Frictional resistance is doubled when the load is doubled.

(3) "The friction made by the same weight will be of equal resistance at the beginning of its movement although the contact may be of different breadths or lengths."

(4) "In friction, every body resists with $\frac{1}{4}$ of its weight, assuming a suitable plane with a polished surface."

(5) "When the inclination of the polished plane enables the body to act with $\frac{1}{4}$ of its weight in the direction of motion, the body is in itself inclined to move downward."

Item (3) is well illustrated by his drawing (Fig. 1) which shows that the same force is required to pull a brick-shaped block along a horizontal plane irrespective of whether it lies on its back, on its side, or stands on end. The drawings of a block on an inclined plane make it evident that da Vinci had a clear idea of the angle of repose and that he probably determined the coefficient of friction as $\frac{1}{4}$ by this means.

The significance of the angle of repose was rediscovered by Antoine Parent[4] in 1704, and further elaborated by the great mathematician, Leonhard Euler[5] (1748) into the form and notation which now appears in our textbooks. The latter also says: "Everyone is agreed that friction depends upon load only and not upon the extent or shape of the surfaces. One would think that, since friction is caused by asperities, the larger the surfaces, the greater the friction because there would be more asperities, but this is not the case." He states that the force of static friction is greater than that of kinetic, and gives an equation for each in terms of the angle of repose.

c. Charles Augustin Coulomb.[6]—The paper, entitled "Théorie des Machines Simples, en Ayant Égard au Frottement de Leur Parties, et à la Roideur des Cordages," which the great physicist, Coulomb, presented before the Royal Academy of Sciences in 1785, is notable for the following findings:

(1) determinations of the factors that control the force of static friction as distinct from those that control that of kinetic friction;

(2) in both cases, Amontons' laws concerning load and contact area (Laws 1 and 2) are valid over the range tested;

(3) in both cases, the force of friction depends upon the nature of the materials in contact and their coatings (Law 4);

(4) the force of static friction depends upon the length of time during which the surfaces have been in contact;

(5) the force of kinetic friction is independent of the velocity (Law 3);

(6) at least part of the frictional force may be ascribed to cohesion of the molecules at the sliding surfaces.

The last suggestion turns out to be a very important one in the light of modern research (Sec. 5). Yet Coulomb summed up the results of his experiments as follows:

"The friction can come only from the engagement of the asperities of the surfaces, and the cohesion should influence it only a little."

His argument is that cohesion should depend upon the number of surface contact points (area); but since frictional force is independent of area over a wide range it is also independent of cohesion. This conclusion is then qualified thus:

"We find, however, that cohesion is not absolutely zero . . . in practice the resistance offered by cohesion can be neglected unless each square foot is loaded by more than 1 quintal" (100 lb).

Coulomb's ideas on the mechanism of friction are similar to those of Amontons. Friction is due to asperities. One piece of polished wood sliding over another is like one hairbrush moving over another; the bristles, or fibers, bend elastically so that they touch without disengaging. Actual disengaging requires the application of a force proportional to the load. When lubricated with tallow, the fibers stick together and so lose a part of their elasticity thus lowering the resistance to motion of the wood.

Since Law 3, that frictional force is independent of the velocity, came to be accepted largely on the authority of Coulomb, it is worth while to know just what he said about it, though it must be admitted his statements are somewhat confusing:

"When the surfaces are large relative to the load, then the friction appears to increase with the speed. But when the surfaces are very small relative to the load, the friction diminishes in proportion as the speed increases . . . it appears that in every practical case one can regard the friction as independent of the speed." (Sec. 42 of Coulomb's paper.)

[4] A. Parent, "Mémoire qui Contient tout ce qui se fait sur les Plans Inclinés," *Mem. de l'Acad. Roy. Sci. de Paris* (1704), p. 173.

[5] L. Euler, "Sur le Frottement des Corps Solides," *Histoire de l'Acad. Roy. Sci. et Belles Lettres* (Berlin 1750), p. 122.

[6] C. A. Coulomb, "Théorie des Machines Simples, En Ayant Égard au Frottement de Leur Parties, et à la Roideur des Cordages," *Mem. de l'Acad. Roy. Sci.* **10**, 161–332 (1785).

He then confirms the last conclusion by showing that the ratio of load to frictional force remains sensibly constant as the sled in his apparatus is accelerated by a falling weight.

Nevertheless, he continues to record observations which indicate that the frictional force does vary with speed. If the surfaces are unlubricated, frictional force increases with speed in the case of (a) elm on elm (velvety to the touch) at low pressures, (b) wood on metals, (c) iron on oak, (d) copper on oak (along the grain) where friction increases in arithmetical progression as speed increases in geometrical progression. (Secs. 51–60.)

If the surfaces are lubricated with tallow, the friction of metals sliding on wood increases with speed (Sec. 72), but if metals slide on other metals, the friction at large loads decreases with increase of speed (Secs. 88–92).

Finally, in his *Essay on the Theory of Friction* (Sec. 95 (4)), Coulomb sums up his observations as follows:

"It is not always that, with wood sliding without lubrication on wood and with metals sliding on metals, the speed has only slight influence upon the friction; but here (Secs. 55 following) the friction increases very sensibly in proportion as the speed is increased."

We must conclude then that Coulomb believed in the dependence of friction upon velocity, in spite of the fact that his range was limited to that near the beginning and that near the end of a 6-ft track along which his sled moved with uniform acceleration (corrected for rotation of the pulley over which the string passed). The accuracy of timing necessary to determine whether or not the acceleration of the sled was truly uniform was beyond the capability of Coulomb's apparatus; hence, it is probably justifiable to conclude that a dependence of friction upon velocity is not *proved* by his experiments, and Law 3 must be considered valid. This, at least, must have been the viewpoint of scientific authorities during the next fifty years, until the careful work of Morin established Law 3 on a firmer foundation.

d. Arthur Morin.[7]—In December, 1831, Morin described before the Royal Academy of Sciences, the first experiments of a series on friction carried out by him over a period of four years. His paper, entitled "Nouvelles Expériences sur le Frottement," was referred to a committee of brilliant men—Poisson, mathematician; Arago, astronomer and physicist; Navier, mechanical engineer—for appraisal, and they made their very laudatory report at the meeting in March, 1832. This report contains the following paragraphs:

"The method of observation employed by Coulomb consisted of making one body slide horizontally over another by means of a weight suspended by a string passing over a pulley. The speed of movement was estimated from the time taken by the slider to traverse the two halves of a 6-ft length, and often the same with only a 4-ft length.

"The results presented in each series of experiments are neither numerous enough nor of sufficient uniformity to give complete certainty to the conclusions.

"Such as they are, they have, nevertheless, permitted their author to state the general laws of the phenomenon, which consist principally as follows: that the frictional resistance of solid bodies is proportional to the pressure exerted by one body upon another and independent of the extent of the surfaces in contact and of the speed of movement.

"It appears also, in consequence of the new observations (of Morin), that these laws are more generally exact and less subject to some exceptions than Coulomb himself had thought."

These conclusions were the ones reached by Morin after summarizing the results of 631 experiments carried out during his four years of work. He extended the range of load to 2,420 lb (1100 kg), of surface area to 3.23 ft^2 (30 dm^2) and of speed to 11.5 ft/sec (3.5 m/sec). His data are numerous and extraordinarily consistent. His observations are made with great care. Hence his conclusions carry conviction. It is to be regretted that his successors have accorded to him less recognition than his work deserves.

Morin changed Coulomb's crude apparatus into a precision instrument by increasing the length of the track to 13.1 ft (4*m*) and by devising accurate automatic recording instruments to measure the motive force applied to the sled at every instant of its travel, and the

[7] A. Morin, "Nouvelles Expériences sur le Frottement," *Inst. de France Acad. Roy. des Sci.* (1831–1834).

nature of the motion which this force communicates to the sled. These instruments, a description of which would be out of place here, were very ingenious and produced accurate records from which the desired magnitudes could be calculated. These beautiful records must be seen in the original paper in order that their perfection may be appreciated.

With regard to Coulomb's theory that resistance to sliding is caused by the bending of "molecular springs" on the surfaces in contact, Morin remarks: "Coulomb, in comparing frictional resistance to that of two brushes whose bristles engage reciprocally, has been led to deduce the hypothesis that the friction of metals on wood ought to increase with speed; and he believes that he has observed this law in his experiments; however, all of our tests show, on the contrary, that in every case observed up to the present, the resistance is independent of the speed, although we have given to our sled speeds which have sometimes reached 3.5 meters per second or more. It is therefore necessary to abandon Coulomb's theory." The validity of Law 3 will be considered further in Sec. 9.

Morin points out, also, that if the fibers behave like "molecular springs" which bend and are relaxed, the amplitude of their oscillations should increase with load; and if the number bent and relaxed in the same time increases with the speed, the vibrations should increase also with the speed. He therefore tested for the existence of oscillations which might be communicated to the track by observing the ripples on the surface of water in a dish placed on the track bed, and the modes of vibration of sand in a long tube parallel either to the length or to the width of the track. In no case were any such vibrations observed; hence he concludes that if any vibrations are produced by sliding, they are not communicated to the support and cause no loss of energy due to friction; therefore frictional resistance must be attributed to some other cause.

3. *Action of Molecular Forces*

Coulomb's theory that the physical cause of frictional phenomena is the "meshing of asperities" (Sec. 2c) leads us to infer that the smoother the sliding surface, the less should be its frictional resistance. That this reasoning is, in many cases, borne out by experience is a matter of common knowledge; but what Coulomb did not suspect is that there is a limit to the smoothness beyond which further polishing *increases* friction,[8] since the more perfect the mating surfaces, the greater is the number of points of contact that lie within the range of the molecular forces that cause cohesion.

In the same year, 1785, both S. Vince[9] in England, and Coulomb in France, suggested independently that molecular attraction may play an important role in causing frictional resistance to sliding motion. If that role were the principal one, they argued, friction should be proportional to the area of contact, which we have seen that it is not (Law 2), but if the meshing of roughnesses is predominant, frictional force should be proportional to the load, as we have seen it is (Law 1). Yet Morin was able to point out the insufficiency of this theory, as many others have done more recently. Thus for the past fifty years there has been much discussion of the importance of molecular forces, and the use of modern experimental techniques has brought about an increased appreciation of the role played by these forces even to the extent of leading to the belief that without them there would be little or no force of friction.

It is interesting, therefore, to find that a modern authority on friction (J. J. Bikerman)[10,11] still advocates the concept that frictional resistance is caused by the meshing of roughnesses; hence, he claims, the correct law of friction must be Law 1, and any phenomena which do not conform to this law cannot be regarded properly as true friction. After making (in 1941) a critical review of recent literature on sliding friction he reached conclusions that may be summarized as follows:

a. Friction cannot be due to the sticking together (welding) of points of contact, since Law 1, including the constant μ, may

[8] H. Ernst and M. E. Merchant, "Surface friction of clean metals," *Proc. Spec. Summer Conference on Friction and Surface Finish* (M.I.T., Cambridge, 1940), pp. 76–101. This is only one of many references that might be cited. See also Sec. 5.

[9] S. Vince, "On the motion of bodies affected by friction," *Phil. Trans.* **75**, 654–668 (1785).

[10] J. J. Bikerman and E. K. Rideal, "A note on the nature of sliding friction," *Phil. Mag.* Ser. 7, **27**, 687 (1939).

[11] J. J. Bikerman, "Friction and adhesion," *Phil. Mag.* Ser. 7, **32**, 67–76 (1941).

be identical for platinum sliding on platinum, platinum on wood, and wood on wood.

b. Friction cannot be due to adhesion, since no adhesion is detected when the normal force is measured (slider lifted off track).

c. Coulomb's law (Law 2) cannot be reconciled with the adhesion hypothesis by assuming the area of actual (instead of apparent) contact is proportional to the load, since this is not in accord with experiments on decreasing load.

d. In certain experiments which seemed to prove that friction is not caused by surface roughnesses it is not the true force of friction that was measured, but instead the "scratch hardness."

e. As long as Law 1 is valid, friction cannot be due to cohesion but to elevation of the load over roughnesses. The absolute values of μ for different surfaces with various degrees of polish are found by measuring the angles of repose, which range from 6° to 27°, corresponding to values of μ from 0.1 to 0.5. This is interpreted to mean that polished surfaces may have irregularities that are inclined to the visible surface at an angle of 6° or more, but when the inclination is greater than 27° the surfaces can no longer be considered smooth and are unsuitable for friction measurements. On the cohesion hypothesis μ might have any value at all.

4. *Conclusions*

In any "practical case," to borrow Coulomb's phrase, the college student is justified in regarding the classical "laws" of friction as valid. This is a viewpoint in agreement with at least one school of contemporary thought, in which these laws are regarded as the expression of all phenomena properly described as "true friction." If this develops in the student's mind a complacent attitude, it is feared he will be unable to maintain it when confronted with the opposing evidence and the more recent theories to be discussed in Part II.

The Quantum-Mechanical Problem of a Particle in Two Adjacent Potential Minima I. Direct Solution

D. S. CARTER* AND G. M. VOLKOFF
The University of British Columbia, Vancouver, British Columbia

THE object of this paper is primarily pedagogic. In introducing a student for the first time to the use of perturbation theory in quantum mechanics it is helpful to have an elementary illustration of the application of the theory to a simple but nontrivial problem which can also be solved exactly, and to use the exact solution for a step-by-step check of the approximations found by perturbation theory. One of few such problems is that of a particle in two adjacent rectangular "potential wells." It is an exactly soluble representative of a class of two adjacent minima problems which are of considerable physical interest in connection with phenomena of metallic conduction,[1] van der Waals forces,[2] the stability of hydrogen-like molecular ions,[2] and the vibration spectra of certain polyatomic molecules.[3]

Exactly soluble cases of two identical one-dimensional adjacent potential wells of various special shapes have been previously treated by Morse and Stückelberg,[4] Dennison and Uhlenbeck,[5] Rosen and Morse,[6] Dushman,[7] and Man-

* Holder of a Studenship from the National Research Council of Canada. Now at Princeton University.

[1] M. F. Manning and M. E. Bell, *Rev. Mod. Physics* **12**, 215 (1940).

[2] S. Dushman, *Elements of quantum mechanics* (Wiley, 1938), pp. 214–218 and references given there.

[3] G. Herzberg, *Infrared and Raman spectra of polyatomic molecules* (Van Nostrand, 1945), pp. 221–224 and references given there.

[4] P. M. Morse and E. G. C. Stückelberg, *Helv. Physica Acta* **4**, 337 (1931).

[5] D. M. Dennison and G. Uhlenbeck, *Physical Rev.* **41**, 313 (1931).

[6] N. Rosen and P. M. Morse, *Physical Rev.* **42**, 210 (1932). See also a remark by D. ter Haar, *Physical Rev.* **70**, 222 (1946), on an error in their treatment.

[7] S. Dushman, *loc. cit.* in ref. 2, considers two identical rectangular wells. His first-order approximation to the split-

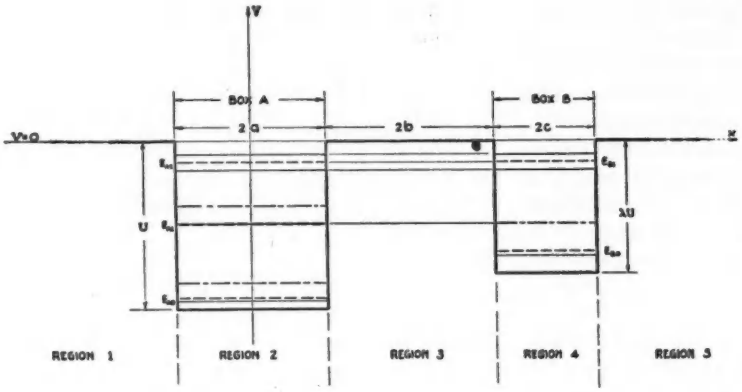


FIG. 1. The two-well potential function $V(x)$ with corresponding energy levels. Solid lines represent levels of the two-well problem, dashed lines represent unperturbed single-well levels, and dot-dashed lines represent "infinite well" levels.

ning.[8] The connection between the exact solution and the perturbation theory approximation for their particular shape of well is discussed by Morse and Stückelberg.[4]

In this paper we consider two not necessarily identical one-dimensional "rectangular wells," a case which does not seem to have been discussed in detail in previous literature. It has been chosen for its mathematical simplicity, and for its pedagogic value in enabling us to draw a clear distinction between two different types of perturbation theory which we designate for brevity as "λ-type" and "ϵ-type." The former, which uses an orthogonal set of unperturbed eigenfunctions, is based on the smallness of the parameter λ which is proportional to the depth of the "perturbing" potential well situated at a constant distance from the "unperturbed" well. This is the usual Schrödinger type of perturbation theory which is discussed in most textbooks on quantum mechanics. The latter is a special type of perturbation theory, illustrated for two identical wells by Morse and Stückelberg,[4] and extended[9] to dissimilar wells in this note. It uses an "almost orthogonal" set of unperturbed wave functions, and is based on the smallness of the parameter ϵ related to the penetrability of the potential barrier separating two wells of fixed depth but variable separation. The pedagogic value of the two-rectangular-well model is further enhanced by the necessity for explicit introduction and use of wave functions of the continuum.

The two rectangular well problem is formulated, and its exact solution in implicit form is obtained by a direct method in Sec. 1. The implicit results are then reduced to explicit form in three special cases: for a single well in Sec. 2, for two wells far apart in Sec. 3, and for one well shallow compared to the other in Sec. 4. In Part II of this paper, which will appear in a subsequent issue of this journal, the results of Sec. 4 are reproduced by standard Schrödinger perturbation theory (λ-type) in Sec. 5, and the results of Sec. 3 are reproduced by perturbation theory using "almost orthogonal" wavefunctions (ϵ-type) in Sec. 6. A discussion of all these formal results is given in Sec. 7. In Sec. 8 the model is applied to the ammonia inversion spectrum.

1. Exact Solution of the Two-Well Problem in Implicit Form

We consider the problem of one-dimensional motion of a particle of mass μ subject to a potential $V(x)$ shown in Fig. 1, and defined by

$$V(x)=\begin{cases}-U & \text{for } |x|<a\\ -\lambda U & \text{for } |x-a-2b-c|<c\\ 0 & \text{for all other } x.\end{cases} \quad (1)$$

We seek the bound energy levels ($E<0$) together with the corresponding eigenfunctions $\phi(x)$ which

ting of degenerate levels lacks a factor $\beta^3 a/(1+\beta a)(\alpha^2+\beta^2)$ (compare his equation (3) on p. 217 with our Eq. (29a) for $c=a$ and $\lambda=1$) since he fails to take into account the phase shift of the single-well eigenfunctions when both wells are present.

[8] M. F. Manning, *J. Chem. Physics* **3**, 136 (1935).

[9] Our discussion follows a line of reasoning outlined by Professor J. R. Oppenheimer in a course of lectures attended by one of us (G.M.V.) at the University of California in 1938.

satisfy both the equation

$$d^2\phi/dx^2+\kappa[E-V(x)]\phi=0 \tag{2}$$

with

$$\kappa=8\pi^2\mu/h^2, \tag{3}$$

and the boundary conditions at infinity (where ϕ must not become infinite), and at the four well boundaries (where ϕ and its first derivative ϕ' must be continuous). The function, defined in the five regions of Fig. 1 by

$$\phi_1=A\cos\alpha(a-\delta)e^{\beta(x+a)} \tag{4a}$$

$$\phi_2=A\cos\alpha(x+\delta) \tag{4b}$$

$$\phi_3=\begin{cases}\dfrac{A\cos\alpha(a+\delta)}{1+\gamma_A}\left[e^{-\beta(x-a)}+\gamma_A e^{\beta(x-a)}\right] & \text{(4c)}\\[2ex] \dfrac{B\cos\alpha'(c+\delta')}{1+\gamma_B}\left[e^{\beta(x-a-2b)}+\gamma_B e^{-\beta(x-a-2b)}\right] & \text{(4d)}\end{cases}$$

$$\phi_4=B\cos\alpha'(x-a-2b-c-\delta') \tag{4e}$$

$$\phi_5=B\cos\alpha'(c-\delta')e^{-\beta(x-a-2b-2c)}, \tag{4f}$$

with

$$\alpha=[\kappa(E+U)]^{\frac{1}{2}} \tag{5a}$$

$$\alpha'=[\kappa(E+\lambda U)]^{\frac{1}{2}} \tag{5b}$$

$$\beta=(-\kappa E)^{\frac{1}{2}} \tag{5c}$$

$$\alpha^2+\beta^2=\kappa U, \quad \alpha'^2+\beta^2=\lambda\kappa U, \tag{5d}$$

satisfies Eq. (1) and the boundary condition at infinity. Expressions (4a) to (4c) and (4d) to (4f), respectively, make up functions which are continuous at the boundaries of wells A and B. The two forms of ϕ_3 have been chosen to emphasize the symmetric way in which the two wells occur in the problem. The six constants A, B, δ, δ', γ_A and γ_B, and the equation for eigenvalues may be determined by the requirements of the continuity of ϕ'/ϕ at the four well boundaries, of the identity of the two expressions (4c) and (4d) for ϕ_3, and of normalization. Continuity of the logarithmic derivative gives

$$\beta/\alpha=\tan\alpha(a-\delta), \tag{6a}$$

$$\beta/\alpha=(1+\gamma_A/1-\gamma_A)\tan\alpha(a+\delta), \tag{6b}$$

and two similar expressions with α, a, δ, and γ_A replaced by α', c, δ' and γ_B. Elimination of δ from Eqs. (6a) and (6b) yields

$$\gamma_A=(\beta^2-\alpha^2+2\alpha\beta\cot2\alpha a)/(\kappa U)$$
$$=(\beta-\alpha\tan\alpha a)(\beta+\alpha\cot\alpha a)/(\kappa U). \tag{7}$$

Similar elimination of δ' yields an expression for γ_B which is simply Eq. (7) with γ_A, α, a and U replaced by γ_B, α', c and λU, respectively. If we now insist that the two forms (4c) and (4d) of ϕ_3 represent the same function, we obtain

$$\gamma_A\gamma_B=\tilde{\epsilon}^2 \tag{8}$$

with

$$\tilde{\epsilon}=e^{-2\beta b}. \tag{9}$$

This is clearly the equation determining the eigenvalues, since in accordance with Eqs. (5) and (7) it is an implicit equation in E and the constants a, b, c, U and λ of the potential function $V(x)$. Both the eigenvalues and the eigenfunctions have thus been implicitly determined. They may be found by numerical methods for any particular case.

2. Specialization to a Single Well

For the application of perturbation theory methods we shall need the unperturbed single well eigenvalues and eigenfunctions for both the discrete and the continuous energy states.

The former may be obtained by specializing the above results by letting $c=0$ or $\lambda=0$. In this case, ϕ_3 represents the eigenfunction for all $x>a$. The boundary condition at infinity now requires that

$$\gamma_A=0, \tag{10}$$

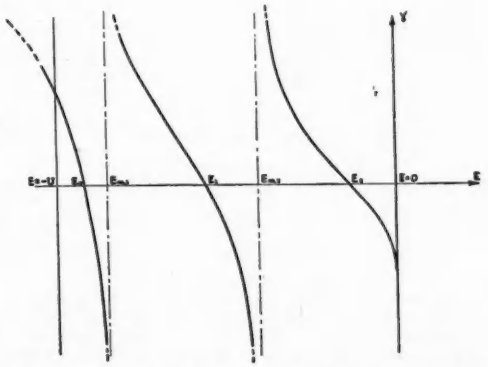


FIG. 2. A rough sketch of $\gamma_A(E)$ for $4<32\mu a^2U/h^2<25/4$. The zeros of $\gamma_A(E)$ define the levels of the single-well problem, and its discontinuities define the "infinite well" levels.

which, with Eqs. (5) and (7), is the condition for eigenvalues.

The trend of γ_A has been roughly sketched in Fig. 2 as a function of E (see Appendix 1 for details of the dependence of $\gamma(E)$ and of its derivative on E) for the case that U lies in the range $4<32\mu a^2U/h^2<25/4$. The value of $\gamma_A(E)$ is real only for $E<0$ and is positive for all $E\leq -U$, approaching $+\infty$ as $E\rightarrow-\infty$. In the region of bound energy levels ($-U<E<0$) the zeros of $\gamma_A(E)$ are separated by infinite discontinuities at the points[10]

$$E_{\infty,n}=n^2h^2/32\mu a^2-U;\quad n=1, 2, 3, \cdots. \tag{11}$$

As E increases, the sign of γ_A changes from negative to positive at each $E_{\infty,n}$ and from positive to negative at each zero. The zeros, when placed in increasing order, are alternately zeros of the first and second factors of γ_A in Eq. (7). The slope of γ_A at each of its zeros E_A is necessarily negative, as may be seen by differentiating expression (7), with respect to the energy E, using Eqs. (5), and using the fact that $\gamma_A(E_A)=0$. The result is

$$\gamma_A'(E_A)=-\kappa^2U(1+\beta a)/2\alpha^2\beta^2. \tag{12}$$

The eigenfunctions ϕ_A corresponding to the bound levels of the single well A may be obtained as follows. We set $\gamma_A=0$ in Eq. (6b), and from Eqs. (6a), (6b) obtain[11]

$$\delta=l\pi/2\alpha, \tag{13}$$

where l is an even or odd integer according as the first or second factor of γ_A in Eq. (7) vanishes. Substitution of Eqs. (10) and (13) into Eqs. (4a) to (4c) shows that the bound state eigenfunctions are alternately multiples of the even and odd functions

$$\begin{aligned}\phi_{A_1}&=\cos\alpha a e^{\beta(x+a)}\\ \phi_{A_2}&=\cos\alpha x\\ \phi_{A_3}&=\cos\alpha a e^{-\beta(x-a)},\end{aligned} \tag{14a}$$

and

$$\begin{aligned}\phi_{A_1}&=-\sin\alpha a e^{\beta(x+a)}\\ \phi_{A_2}&=\sin\alpha x\\ \phi_{A_3}&=\sin\alpha a e^{-\beta(x-a)},\end{aligned} \tag{14b}$$

according as the first or second factor of γ_A vanishes.

In both cases it may be shown that the normalizing factor C_A is determined by

$$\frac{1}{C_A^2}=\int_{-\infty}^{+\infty}\phi_A^2dx=a+1/\beta=\frac{2\alpha^2\beta}{\kappa^2U}[-\gamma_A'(E_A)], \tag{15}$$

where the prime denotes differentiation with respect to E. For later use we observe, with the aid of Eq. (7), that for the even functions the following relations hold:

$$\cos^2\alpha a=\alpha^2/(\alpha^2+\beta^2)=\alpha^2/\kappa U; \tag{16a}$$

and for the odd functions

$$\sin^2\alpha a=\alpha^2/(\alpha^2+\beta^2)=\alpha^2/\kappa U. \tag{16b}$$

The single-well eigenfunctions for the continuum of unbound states ($E>0$), which are also needed for perturbation theory calculations, are obtained as follows. Setting

$$k=(\kappa E)^{\frac{1}{2}}>0\quad\text{for}\quad E>0, \tag{17}$$

we find that outside the well A, ϕ no longer has the exponential form of expressions (4a), (4c), but that it has the oscillatory form $A\cos k(x+\delta)$. It is clear that Eq. (2) is satisfied for every $E>0$ by the following linearly independent pair of functions φ_k^+ and φ_k^-, of which the first is even and the second is odd:

$$\begin{aligned}\varphi_{k1}^{\pm}&=\pm\pi^{-\frac{1}{2}}\cos k(x-\delta^{\pm})\\ \varphi_{k2}^{+}&=\pi^{-\frac{1}{2}}\frac{\cos k(a+\delta^{+})}{\cos\alpha a}\cos\alpha x\\ \varphi_{k2}^{-}&=\pi^{-\frac{1}{2}}\frac{\cos k(a+\delta^{-})}{\sin\alpha a}\sin\alpha x\\ \varphi_{k3}^{\pm}&=\pi^{-\frac{1}{2}}\cos k(x+\delta^{\pm}).\end{aligned} \tag{18}$$

The functions $\varphi_k^{\pm}$ defined by Eq. (18) are already continuous at the well boundaries, and satisfy the boundary condition at infinity. The further requirement of the continuity of their logarithmic derivatives determines δ^+ and δ^- by the equations

$$\begin{aligned}k\tan k(a+\delta^{+})&=\alpha\tan\alpha a\\ k\tan k(a+\delta^{-})&=-\alpha\cot\alpha a.\end{aligned} \tag{19}$$

Each level of the continuum is thus seen to be twofold degenerate.

[10] As $U\rightarrow\infty$, each zero approaches the $E_{\infty,n}$ which lies immediately above it. Thus the $E_{\infty,n}$ may be thought of as the levels of an infinitely deep box of the same width with base at $-U$.

[11] To avoid extra notation we shall use α and β to denote the functions of E defined by Eqs. (5), and also the special values which these functions assume corresponding to eigenvalues of E.

The factor $1/(\pi)^{\frac{1}{2}}$ in Eqs. (18) has been chosen to make the $\varphi_k^{\pm}$ normalized to Dirac's delta-functions (see Appendix 2).

The eigenvalues and eigenfunctions belonging to the single well B may be obtained from the above discussion by replacing a, U, α, x, γ_A, δ, respectively, by c, λU, α', $x-a-2b-c$, γ_B, δ', and leaving all other quantities unchanged. We shall denote by ϕ_B the bound state eigenfunctions found in this way from Eqs. (14).

3. Specialization to the Case of Two Wells Far Apart

We return to the results of the first section and examine them in the case that the distance $2b$ between the two wells is large. For arbitrarily large b, $\tilde{\epsilon}$ of Eq. (9) is correspondingly small, and the value of $\gamma_A\gamma_B$ in Eq. (8) is arbitrarily close to zero. It follows that every two-well level E is arbitrarily close to a level E_0 of one or the other of the single wells A and B (a zero of γ_A or γ_B). Indeed, for sufficiently large b, every such E may be approximated by a solution of the equation

$$[\gamma_A(E_0)+(E-E_0)\gamma_A'(E_0)] \times[\gamma_B(E_0)+(E-E_0)\gamma_B'(E_0)]=\epsilon^2 \quad (20)$$

where γ_A' and γ_B' are derivatives with respect to E, and

$$\epsilon=\tilde{\epsilon}(E_0)=e^{-2b\sqrt{(-\kappa E_0)}}. \quad (21)$$

Since both Eqs. (8) and (20) are symmetric in γ_A and γ_B, we need discuss only the case in which $E_0=E_A$ is a level of well A. We can then obtain results for the case $E_0=E_B$ by symmetry. Substituting $\gamma_A(E_A)=0$ in Eq. (20), and solving for that root $E-E_A$ which approaches zero as $\epsilon\to 0$, we obtain

$$E-E_A=\frac{\gamma_B(E_A)}{2\gamma_B'(E_A)}[(1+X_A)^{\frac{1}{2}}-1] \quad (22)$$

$$\approx\frac{\epsilon^2}{\gamma_A'(E_A)\gamma_B(E_A)} \quad \text{for} \quad |X_A|\ll 1 \quad (23)$$

$$\approx\pm\epsilon[\gamma_A'(E_A)\gamma_B'(E_A)]^{-\frac{1}{2}} \quad \text{for} \quad X_A\gg 1, \quad (24)$$

where

$$X_A=[4\epsilon^2\gamma_B'(E_A)]/[\gamma_A'(E_A)\gamma_B^2(E_A)]. \quad (25)$$

Since $\gamma_A'(E_A)$ is always negative (Eq. (12)), X_A agrees in sign with $-\gamma_B'(E_A)$, and accordingly $E-E_A$ agrees in sign with $-\gamma_B(E_A)$. This tells us which sign must be chosen in Eq. (24). Further, if E_B is that level of well B which is not separated from E_A by an infinite discontinuity of $\gamma_B(E)$, and which we call for brevity the level of well B that is "closest to E_A," $-\gamma_B(E_A)$ has the same sign as E_A-E_B. Hence E lies on the side of E_A opposite to E_B; *i.e.*, E_A is "repelled" by the "closest" level of well B.

There are two special cases in which the last remarks are not strictly applicable: when E_A coincides exactly with a level E_B (the "closest" level) of well B; and when E_A coincides with an infinite discontinuity of $\gamma_B(E)$ (E_A is "equally close" to two levels of well B). In the former case we find by substituting $\gamma_A(E_A)=\gamma_B(E_A)=0$ into Eq. (20), that there are two solutions for $E-E_A$ given by expression (24) with both choices of sign. In the latter case we find by substituting $\gamma_A(E_A)=0$ and $\gamma_B(E_A)=\infty$ into the exact Eq. (8), that E_A is a level of the two-well system for all values of b and ϵ, *i.e.*, $E-E_A\equiv 0$.

Those cases in which Eqs. (23) and (24) are valid are of special interest. In the former case, which we call "completely nondegenerate," two conditions are satisfied: (a) E_A does not coincide with any level of well B, so that $\gamma_B(E_A)\neq 0$, and (b) the wells are far enough apart and ϵ is sufficiently small that $|X_A|\ll 1$. In the latter case, which we call "degenerate,"[12] either (a) E_A coincides exactly with a level of well B ("exact degeneracy") or (b) E_A is sufficiently close to a level of well B ($|\gamma_B(E_A)|$ is so much smaller than unity)[13] and ϵ is sufficiently large that $X_A\gg 1$ ("approximate degeneracy").

The following interpretation in terms of level spacings of conditions of "complete nondegeneracy" and of "degeneracy," respectively, under which expressions (23) and (24) are valid is instructive. Let us consider two single well levels E_A and E_B which lie sufficiently closely together

[12] In this use of the word "degenerate" with reference to a case in which there is a simultaneous level of two *different* problems we differ from customary usage. Ordinarily the word denotes the existence of more than one linearly independent eigenfunction belonging to a single level of the *same* problem, as in the case of Eqs. (18) above.

[13] It is shown in Appendix 1 that $\gamma_B'(E_A)$ is negative whever $|\gamma_B(E)|<1$. Hence the condition $|\gamma_B(E_A)|<1$ ensures that X_A shall be positive.

that we may write, approximately,

$$\gamma_B(E_B)\equiv 0\approx\gamma_B(E_A)+(E_B-E_A)\gamma_B'(E_A): \quad (26)$$

and consequently

$$E_A-E_B\approx\gamma_B(E_A)/\gamma_B'(E_A). \quad (27)$$

Substitution of Eq. (27) into Eq. (25) gives

$$X_A\approx 4\epsilon^2/[\gamma_A'(E_A)\gamma_B'(E_A)(E_A-E_B)^2]$$
$$\approx 4\epsilon^2/[\gamma_A'(E_A)\gamma_B(E_A)(E_A-E_B)]. \quad (28a)$$

A further substitution of Eq. (23) into the third term of Eq. (28a) gives, for the completely nondegenerate case,

$$|X_A|\approx 4\left|\frac{E-E_A}{E_A-E_B}\right|\ll 1. \quad (28b)$$

A similar substitution of Eq. (24) into the second term of Eq. (28a) gives, for the degenerate case,

$$X_A\approx 4(E-E_A)^2/(E_A-E_B)^2\gg 1. \quad (28c)$$

Thus the conditions for complete nondegeneracy, or for degeneracy, of a pair of neighboring single-well levels belonging to different wells are, respectively, that the wells must be far enough apart, or close enough together, that the shift of a level E_A due to the presence of the other well B must, in the former case, be much smaller, and in the latter case much larger, than the original small level separation E_A-E_B. This last result shows that if in expression (24), $\gamma_B'(E_A)$ is replaced by $\gamma_B'(E_B)$, the resultant expression will still give the level shift correctly to terms of order ϵ.

Stated in words, the above results for $E-E_A$ and their analogs for $E-E_B$ imply the following: The system of two wells at a sufficiently large but finite distance apart has a level E corresponding to each level E_0 of the two wells considered independently of each other (at infinite separation). For each E_0, there is a range of distances between the wells (a range of values of ϵ) over which our approximations for $E-E_0$ are valid. Within this range, the form of $E-E_0$ will vary in accordance with the size of ϵ and the distance of E_0 from the "closest" level of the opposite well. Thus, if E_0 happens to coincide exactly with a level of the other well (either accidental degeneracy, when two levels of dissimilar wells coincide by chance, or complete degeneracy, when all levels of two identical wells coincide), Eq. (24) tells us that the two-well levels corresponding to E_0 lie one above, and the other equally far (to terms of order ϵ) below E_0, their splitting always being of order ϵ. If E_0 lies very close to a level of the other well, then for very small values of ϵ (large separations of the wells) the case will be nondegenerate, and Eq. (23) tells us that the level shift will be of order ϵ^2. As ϵ is increased (the wells are brought closer together), however, complete nondegeneracy will no longer exist, and the shift will no longer be of order ϵ^2. Eventually, for the largest permissible values of ϵ, approximate degeneracy will set in, and the shift will become of order ϵ. Finally, if E_0 is not near a level of the opposite well, the case will be completely nondegenerate for the smallest values of ϵ, and although it may not necessarily remain so for larger values of ϵ, it will never become approximately degenerate.

The above results are qualitatively illustrated by the levels sketched in Fig. 1. The lowest level of the two box system lies below the lowest level E_{A0} of well A, since it is necessarily "repelled" by the lowest level E_{B0} of well B. Similarly, the second level of the two-well system lies below E_{B0} which is "repelled" by E_{A1}. The third level coincides with E_{A1}, since the latter coincides with $E_{\infty,1}$ of well B. Of the next pair of levels, one lies above, and the other below, the accidentally coincident levels E_{A2} and E_{B1}.

Making use of Eqs. (7) and (12) and their analogs for well B we rewrite Eqs. (24) and (23) in explicit form to facilitate comparison with Dushman[7] and with the results of perturbation theory. Thus for degenerate levels (using $\gamma_B'(E_A)\approx\gamma_B'(E_B)$ to the desired order of accuracy) we obtain

$$E-E_A=\frac{\pm 2\epsilon\alpha\alpha'\beta^2}{\kappa^2 U[\lambda(1+\beta a)(1+\beta c)]^{\frac{1}{2}}} \quad (29a)$$

where both signs are used for exact degeneracy, while for approximate degeneracy the sign is chosen in agreement with that of E_A-E_B. For completely nondegenerate levels the corresponding relation is

$$E-E_A=\frac{-2\epsilon^2\lambda\alpha^2\beta^2}{\kappa(1+\beta a)(\beta^2-\alpha'^2+2\alpha'\beta\cot 2\alpha' c)}. \quad (29b)$$

We now turn to a discussion of the eigenfunctions of the two-well problem. Again because of symmetry we need discuss only those cases in which the two-well level E lies close to a level E_A of well A. In these cases $\gamma_A(E)$ is almost zero, the values of δ determined by Eqs. (6) differ little from those given by Eq. (13) and the first three of Eqs. (4) differ little from Eqs. (14). Thus in the neighborhood of the well to which the unperturbed level belongs, the two-well eigenfunction differs little from the corresponding single-well eigenfunction.

To determine the nature of the perturbed eigenfunction near well B we find the ratio $|B/A|$ of the amplitudes of the oscillatory parts of the eigenfunction inside the two wells. First we equate expressions (4c) and (4d) at $x=a$, and eliminate $\tilde{\epsilon}$ by means of Eq. (8). Then we simplify by means of Eqs. (5d), (6) and (7) to obtain

$$\left|\frac{B}{A}\right| = \left(\left|\frac{\gamma_A}{\gamma_B}\right|\right)^{\frac{1}{2}} \left|\frac{(1+\gamma_B)\cos\alpha(a+\delta)}{(1+\gamma_A)\cos\alpha'(c+\delta')}\right| = \frac{1}{(\lambda)^{\frac{1}{2}}}\left(\left|\frac{\gamma_A}{\gamma_B}\right|\right)^{\frac{1}{2}} \left|\frac{\sin 2\alpha a}{\sin 2\alpha' c}\right|, \quad (30)$$

where all functions should be evaluated at the actual eigenvalue E of the two-well problem.

In the completely nondegenerate case we may use, as an approximation, Eq. (8) with $\tilde{\epsilon}$ replaced by ϵ to eliminate γ_A, and setting $E \approx E_A$ obtain

$$\left|\frac{B}{A}\right| \approx \frac{\epsilon}{(\lambda)^{\frac{1}{2}}|\gamma_B(E_A)|}\left|\frac{\sin\alpha a}{\sin\alpha' c}\right| \quad (31)$$

which is correct to order ϵ. This shows that the ratio of the amplitude of the eigenfunction in the perturbing well to that in the perturbed well is of order ϵ, and hence the relative probability of finding the particle near the perturbing well is only of order ϵ^2 compared with that of finding it near the well to which the original level E_A belongs.

In the degenerate case we obtain with the aid of the relations $\gamma_A(E_A)=\gamma_B(E_B)=0$,

$$\gamma_A(E) \approx (E-E_A)\gamma_A'(E_A) \quad (32)$$

$$\gamma_B(E) \approx (E-E_B)\gamma_B'(E_B). \quad (33)$$

Further, since expression (7) and its analog for well B are both nearly zero we obtain with the aid of Eqs. (5d),

$$|\sin 2\alpha a| \approx 2\alpha\beta/(\kappa U); \quad |\sin 2\alpha' c| \approx 2\alpha'\beta'/(\lambda\kappa U), \quad (34)$$

where α and β belong to E_A, and α', β' belong to E_B. Substituting Eqs. (32), (33) and (34) into Eq. (30), and then substituting Eq. (15) and its analog into the result, we obtain

$$\left|\frac{B}{A}\right| \approx \frac{1}{(\lambda)^{\frac{1}{2}}}\left|\frac{\sin 2\alpha a}{\sin 2\alpha' c}\right|\left(\frac{\gamma_A'(E_A)(E-E_A)}{\gamma_B'(E_B)(E-E_B)}\right)^{\frac{1}{2}} \approx \frac{C_B}{C_A}\left(\frac{\beta(E-E_A)}{\beta'(E-E_B)}\right)^{\frac{1}{2}}, \quad (35)$$

where C_A and C_B are the normalizing factors of ϕ_A and ϕ_B defined by Eq. (15) and its analog. In the case of *exact* degeneracy, when $E_A=E_B$, and $\beta=\beta'$, Eq. (35) reduces for *all* values of ϵ to

$$|B/A| \approx C_B/C_A \quad (36)$$

which shows that no matter how far apart the wells are, and how small the wave function is in the region between the two boxes, the normalized wave functions ϕ for the two levels of the two-well problem, into which the original degenerate level splits, are given by

$$\phi \approx (C_A\phi_A \pm C_B\phi_B)/\sqrt{2}. \quad (37)$$

Further, the ratio of the probability of finding the particle in the neighborhood of one well to that of finding it in the neighborhood of the other is unity. For the approximately degenerate case ($\beta\approx\beta'$ and $E-E_A\approx E-E_B$) Eq. (35) shows that the ratio of these two probabilities is given by

$$\beta(E-E_A)/\beta'(E-E_B) \quad (38)$$

which differs from unity at most by terms of order ϵ. As the separation of the two wells is increased and $\epsilon\to 0$, $|E-E_A|$ becomes smaller than $|E-E_B|$, so that the relative probability of finding the particle near well B as compared with well A diminishes, and the two-well level E which corresponds to E_A takes on more and more of the characteristics of E_A.

4. Specialization to the Case in Which One of the Wells Is Shallow

We again return to the results of the first section, and examine them in the case that b is arbitrary, but fixed, and λ is small (well B is shallow). We obtain an explicit approximate expression for the eigenvalues by considering the fact that Eq. (8) defines E as a many-valued function of λ. It is clear that for every level E_A of the single well A there is a branch of this function which approaches E_A as $\lambda \to 0$. But each of these branches may be expanded in a series of the form

$$E = E_A + \lambda(dE/d\lambda) + \tfrac{1}{2}\lambda^2(d^2E/d\lambda^2) + \cdots, \quad (39)$$

where the derivatives are all evaluated at $\lambda = 0$ and $E = E_A$. The range of applicability of these series will be discussed later.

The coefficients of λ and λ^2 in the above expansion have been determined from Eq. (8) by partial differentiation (see Appendix 3) in order to compare the direct result of Eq. (39) with that of perturbation theory. They are

$$E_1 = (dE/d\lambda)_{\lambda=0} = \epsilon^2\alpha^2(\epsilon^{-4\beta c} - 1)/2\kappa(1+\beta a) \quad (40a)$$

and

$$E_2 = \tfrac{1}{2}(d^2E/d\lambda^2)_{\lambda=0} = \epsilon^2\alpha^2 U[(1+4\beta c)e^{-4\beta c} - 1]/4\beta^2(1+\beta a) + E_1{}^2(\bar{\epsilon}^{-2}d\bar{\epsilon}^2/dE - \gamma_B'/\gamma_B - \tfrac{1}{2}\gamma_A''/\gamma_A')_{\lambda=0}, \quad (40b)$$

where E_1 denotes expression (40a), the primes denote differentiation with respect to E, and all the functions are evaluated at $E = E_A$.

Appendix 1. Some Properties of $\gamma(E)$ and $\gamma'(E) \equiv d\gamma(E)/dE$ for $E < 0$

Setting $M^2 = a^2\kappa U$ and $x = \alpha a$ in Eqs. (7) and (5) we obtain:

$$\gamma(E) = (M^2 - 2x^2 + 2x(M^2 - x^2)^{\frac{1}{2}} \cot 2x)/M^2. \quad (A1)$$

Differentiating and combining terms we find

$$\gamma'(E) \equiv d\gamma(E)/dE = (\kappa a^2/2x)(d\gamma/dx) = -\frac{2\kappa a^2}{M}\left[1 + (M^2 - x^2)^{\frac{1}{2}} \csc^2 2x \times \left\{1 + \frac{\sin 4x}{4x}\left(\frac{2x^2 - M^2}{M^2 - x^2}\right)\right\}\right]. \quad (A2)$$

For $E < -U$, x is purely imaginary, $\gamma(E)$ is positive, and $\gamma'(E)$ is negative, as may be seen from Eqs. (A1) and (A2). For $-U < E < 0$, x lies between 0 and M. In this range $\gamma(E)$ has infinite discontinuities at the points $x_n = n\pi/2$, where n is any positive integer. At $E = 0$, $x = M$ and $\gamma(E)$ assumes the value $\gamma(0) = -1$.

We now investigate the sign of $\gamma'(E)$ in the range $-U < E < 0$, or $0 < x < M$. The expression

$$(2x^2 - M^2)/(M^2 - x^2) \quad (A3)$$

increases monotonically from -1 to 0 to $+\infty$ as x increases from 0 to $M/\sqrt{2}$ to M. Since $|\sin 4x/4x| < 1$, we see that $\gamma'(E)$ is necessarily negative for $0 < x \leq M/\sqrt{2}$. For x in the range $M/\sqrt{2} < x < M$, expression (A3) is positive, and hence the quantity within the curly brackets in expression (A2) will be negative if, and only if,

$$\sin 4x/4x < -(M^2 - x^2)/(2x^2 - M^2). \quad (A4)$$

A graph of the functions on the two sides of this inequality, and an investigation of their derivatives, shows that if M is such that

$$2n\pi < 4M < (2n+1)\pi, \quad (A5)$$

where n is a positive integer or zero, the inequality (A4) cannot hold for any value of x in the range $M/\sqrt{2} < x < M$. If M is such that

$$(2n+1)\pi < 4M < (2n+2)\pi, \quad (A6)$$

where n is a positive integer or zero, there will be an intersection of the two curves at a point x_0 such that

$$(2n+1)\pi < 4x_0 < 4M < (2n+2)\pi. \quad (A7)$$

Thus when M satisfies condition (A5), $\gamma'(E)$ is negative everywhere; but when M satisfies condition (A6), $\gamma'(E)$ is negative everywhere except possibly within that range of x defined by

$$(2n+1)\pi < 4x_0 < 4x < 4M < (2n+2)\pi. \quad (A8)$$

The zeros of $\gamma(E)$ correspond to those values of x for which the functions $(2x^2 - M^2)/(M^2 - x^2)^{\frac{1}{2}}$ and $2x/\tan 2x$ are equal. Graphs of these functions show that there is one zero in each subinterval into which the interval $0 < x < M$ is divided by the points of infinite discontinuity $x_n = n\pi/2$. Since the former function is positive in the range $0 < x < M/\sqrt{2}$, those zeros of $\gamma(E)$ which lie below $M/\sqrt{2}$ necessarily lie in the intervals $x_n - \pi/4 < x < x_n$, while those which lie between $M/\sqrt{2}$ and M necessarily lie in the intervals $x_n < x < x_n + \pi/4$. Comparing this with condition (A8) we see that the derivative $\gamma'(E)$ can possibly become positive only for values of x greater than the highest zero of $\gamma(E)$.

The value which $\gamma'(E)$ approaches as $E \to 0$ may be obtained by setting $x = M - u$ in Eq. (A2), and letting $u \to 0$. We obtain

$$\gamma' \to -\frac{\kappa a^2 \csc^2 2M}{2(2Mu)^{\frac{1}{2}}} \sin 4M. \quad (A9)$$

This is negative or positive according as M satisfies condition (A5) or (A6). In the former case the graph of $\gamma(E)$ is represented by the curve of Fig. 2. In the latter case the character of the graph remains the same for energies below the highest zero; but above the highest zero the graph first drops below -1, and then approaches the value -1 with positively infinite slope as $E \to 0$.

Appendix 2. Normalization of the Eigenfunctions of the Continuum

The λ-type perturbation theory results summarized in Sec. 5 are obtained on the condition that the unperturbed eigenfunctions of the discrete states of the one-well problem are normalized to unity, while the eigenfunctions belonging to the states in the continuum are normalized to Dirac delta-functions by the requirement

$$\int_{-\infty}^{+\infty}\phi_k(x)\phi_{k'}(x)dx=\delta(k-k'), \tag{A10}$$

where the ϕ_k are both odd or both even eigenfunctions defined by Eq. (18). Condition (A10) is equivalent to

$$2\int_0^{\infty}\phi_k(x)\int_{k-\Delta}^{k+\Delta}\phi_{k'}(x)dk'dx=1, \tag{A11}$$

where Δ is positive but may be arbitrarily small. Taking limits of both sides of Eq. (A11) as $\Delta\to 0$, we see that no finite contribution to the limit of the integral over x can come from any finite range of integration, so that Eq. (A11) may be replaced by

$$2\lim_{\Delta\to 0}\int_{x_0}^{\infty}\phi_k(x)\int_{k-\Delta}^{k+\Delta}\phi_{k'}(x)dk'dx$$
$$=2\lim_{\Delta\to 0}\int_0^{\infty}\phi_k(y+x_0)\int_{k-\Delta}^{k+\Delta}\phi_{k'}(y+x_0)dk'dy=1, \tag{A12}$$

where x_0 is an arbitrarily large, but finite number. Thus only the *asymptotic* form of $\phi_k(x)$ is important in determining the normalizing factor. It is given from Eq. (18) by

$$\phi_k(x)=C_k\cos k(x+\delta)\quad\text{for}\quad x>a, \tag{A13}$$

where δ depends on k. If we choose $x_0>a$, and such that

$$\cos k(x_0+\delta)=1, \tag{A14}$$

we may rewrite Eq. (A12) in the form

$$2\lim_{\Delta\to 0}\int_0^{\infty}C_k\cos ky\int_{k-\Delta}^{k+\Delta}C_{k'}[\cos k'(x_0+\delta')\cos k'y$$
$$-\sin k'(x_0+\delta')\sin k'y]dk'dx=1, \tag{A15}$$

where δ' is the value of δ for $k=k'$. It may be shown that since $\cos k'(x_0+\delta')=1$ and $\sin k'(x_0+\delta')=0$ for $k'=k$, Eq. (A15) may again be rewritten

$$2\lim_{\Delta\to 0}\int_0^{\infty}C_k\cos ky\int_{k-\Delta}^{k+\Delta}C_{k'}\cos k'ydk'dy=1. \tag{A16}$$

Finally, it may be shown that $C_k\equiv 1/(\pi)^{\frac{1}{2}}$ satisfies condition (A16).

Appendix 3. Derivation of Eqs. (40a) and (40b)

From Eqs. (8), (5), (21), and the analog of Eq. (7) we have

$$E'\equiv\frac{dE}{d\lambda}=-\frac{\partial}{\partial\lambda}(\gamma_A\gamma_B-\bar{\epsilon}^2)\Big/\frac{\partial}{\partial E}(\gamma_A\gamma_B-\bar{\epsilon}^2)$$
$$=\bar{\epsilon}^2\left(\frac{\gamma_B}{\lambda}-\frac{\partial\alpha'}{\partial\lambda}\frac{\partial\gamma_B}{\partial\alpha'}\right)\Big/\left(\bar{\epsilon}^2\gamma_B'+\gamma_B{}^2\gamma_A'-\gamma_B\frac{d\bar{\epsilon}^2}{dE}\right), \tag{A17}$$

there γ_A' and γ_B' are derivatives with respect to E. Setting $\lambda=0$ and $E=E_A$ we obtain

$$E_1=(E')_{\lambda=0}=(\bar{\epsilon}^2/\lambda\gamma_B\gamma_A')_{\lambda=0}. \tag{A18}$$

Setting $\lambda=0$, $E=E_A$, and $\alpha'=i\beta$ (from Eqs. (5b) and (5c) with $\lambda=0$) in the analog of Eq. (7), we obtain

$$(\lambda\gamma_B)_{\lambda=0}=4\beta^2/\kappa U(1-e^{-4\beta c}). \tag{A19}$$

Substitution of Eq. (A19) into Eq. (A18), and application of Eqs. (12) and (21) leads, for $\lambda=0$, $E=E_A$, directly to Eq. (40a).

To obtain Eq. (40b) we differentiate Eq. (A17);

$$E''\equiv\frac{d^2E}{d\lambda^2}=\left(\frac{\partial E'}{\partial\lambda}\right)_{\alpha'}+\left(\frac{\partial E'}{\partial\alpha'}\right)_{\lambda}\frac{\partial\alpha'}{\partial\lambda}+E'\frac{\partial E'}{\partial E}. \tag{A20}$$

For $\lambda=0$ we find

$$\left(\frac{\partial E'}{\partial\lambda}\right)_{\alpha'}=\frac{\bar{\epsilon}^2}{\lambda^2\gamma_B{}^2\gamma_A'{}^2}\left(\gamma_B\frac{d\bar{\epsilon}^2}{dE}-\bar{\epsilon}^2\gamma_B'\right)$$
$$-\frac{\bar{\epsilon}^2}{\lambda\gamma_B{}^2\gamma_A'}\cdot\frac{\partial\gamma_B}{\partial\alpha'}\cdot\frac{\partial\alpha'}{\partial\lambda} \tag{A21}$$

$$\left(\frac{\partial E'}{\partial\alpha'}\right)_{\lambda}=-\frac{\bar{\epsilon}^2}{\lambda\gamma_B{}^2\gamma_A'}\cdot\frac{\partial\gamma_B}{\partial\alpha'} \tag{A22}$$

$$\frac{\partial E'}{\partial E}=\frac{1}{\lambda\gamma_B\gamma_A'}\cdot\frac{d\bar{\epsilon}^2}{dE}-\frac{\bar{\epsilon}^2}{\lambda\gamma_B{}^2\gamma_A'{}^2}(\gamma_A'\gamma_B'+\gamma_B\gamma_A''). \tag{A23}$$

Substituting into Eq. (A20) we find, with the help of Eq. (A18), that

$$(E'')_{\lambda=0}=2E_2=-2E_1\left[\frac{1}{\gamma_B}\cdot\frac{\partial\gamma_B}{\partial\alpha'}\cdot\frac{\partial\alpha'}{\partial\lambda}\right]_{\lambda=0}$$
$$+E_1{}^2\left[2\left(\frac{1}{\bar{\epsilon}^2}\frac{d\bar{\epsilon}^2}{dE}-\frac{\gamma_B'}{\gamma_B}\right)-\frac{\gamma_A''}{\gamma_A'}\right]_{\lambda=0}. \tag{A24}$$

Finally, we find $\partial\alpha'/\partial\lambda$ and $\partial\gamma_B/\partial\alpha'$ from Eq. (5b) and the analog of Eq. (7), respectively, set $\alpha'=i\beta$, and use Eqs. (40a) and (A24) to obtain Eq. (40b).

Liquid helium is a starting point for all work in the region within only a degree or so of absolute zero. The importance of investigations at this temperature is that the atoms have been robbed of almost all their thermal motion. In consequence a whole new range of physical properties can be studied, which is obscured by the irregular motion of the atoms at higher temperatures. Just as the anxious photographer tells his audience to keep quite still while he takes the cap off his camera, so the physicist likes the atoms to keep quiet while trying to discover some of their more delicate and small-scale properties.—W. L. BRAGG.

The Exact Representation of a Series of Points by a Polynomial in Power Series Form

RAYMOND T. BIRGE
University of California, Berkeley, California

A FREQUENTLY occurring problem in physical science is the exact representation of a series of points $y_i = f(a_i)$ by a polynomial, or rational integral function as it is more logically designated. All ordinary interpolation formulas correspond to such a representation, where the polynomial is in factorial form. Thus the Gregory-Newton interpolation formula—the oldest known interpolation formula—expresses the result explicitly in terms of leading differences. The Newton-Stirling, Newton-Bessel, and Newton-Gauss formulas employ central differences. All of these formulas apply only to the case of equally spaced abscissa values. If the arguments are unequally spaced, we are forced to use "divided differences," and the corresponding equation, also in factorial form, is known as Newton's equation.

But the physical scientist often prefers to obtain the desired polynomial in power series form. A commonly used method for getting such a result, in the case of equally spaced arguments, is described in Section B of a recent paper by the writer.[1] In that paper the statement is made that the method presented in Section B is "the most rapid and convenient one for setting up a polynomial in power-series form which exactly satisfies a given set of equally spaced data."

The purpose of the present paper is to show that the foregoing statement is far from correct. As the writer has recently discovered, the method of *divided differences with repeated arguments* is a better one to use for the case of equally spaced arguments, and it has the very great advantage that it is applicable also to the entirely general case of unequally spaced arguments. But its most important advantage lies in the fact that no explicit formulas are required. All calculations may be carried out directly by means of a properly extrapolated table of divided differences.

Although this method of divided differences with repeated arguments is not new, no emphasis is given to its possibilities and importance in the few places where I have found it discussed,[2,3] and I have never seen it used in any scientific paper. In order to bring out as clearly as possible the advantages of the method, I give first a brief outline of the method discussed in Section B of reference 1. Further details, together with a fully worked out illustrative problem, may be found there.

One first sets up a polynomial in factorial form in terms of either leading or central differences. This expression is then multiplied out, in order to collect all terms in each power of the independent variable. An interpolation formula in terms of central differences is much better, for this purpose, than one in terms of leading differences. When, for instance, the Newton-Stirling formula is thus used, the resulting power series, carried only through the third central difference, appears as

$$u_\epsilon = u_0 + \epsilon[\mu\delta u_0 - (\mu\delta^3 u_0/6)] + \epsilon^2(\delta^2 u_0/2) + \epsilon^3(\mu\delta^3 u_0/6). \quad (1)$$

There is no general form for such an equation. Hence it cannot be extended to higher differences without starting again with the interpolation formula (carried to a sufficient number of terms) and again multiplying out. The resulting expression, carried through the eighth central difference, is given in full on page 304 of reference 1, and, except for the omission of the seventh and eighth degree terms, on page 65 of reference 2.

In Eq. (1) ϵ is a special abscissa corresponding to unit interval and zero value at the center of the data. If the original abscissa is x, where $x = \epsilon h + a = e + a$, one first converts the $f(\epsilon)$ of Eq. (1) to $f(e)$, by dividing the coefficient of each ϵ^i by h^i and replacing ϵ by e. Finally, one converts from $f(e)$, or $f(x-a)$, to $f(x)$ by a Horner shift (synthetic division) of magnitude $-a$.

When the abscissa values are unequally spaced, and divided differences must be used in setting up a polynomial in factorial form (Newton's

[1] R. T. Birge, *Rev. Mod. Physics* **19**, 298–360 (1947), see pp. 303–305.

[2] E. T. Whittaker and G. Robinson, *The calculus of observations* (Blackie and Son, 1924), pp. 27–28.

[3] L. M. Milne-Thomson, *The calculus of finite differences* (Macmillan, 1933), pp. 12–14.

equation), there is obviously an equation, corresponding in a general way to Eq. (1), by which Newton's equation may be transformed to a power series form. But in the case of divided differences, the original abscissa values (i.e., values of x) already appear explicitly in Newton's equation, and the resulting form appears as a power series in $(x-a_0)$. Hence only a Horner shift of magnitude $-a_0$ is then necessary to convert to the desired power series in x.

I have never seen in print such a transformation of Newton's equation to a power series form, but Dr. D. C. Kalbfell, while a graduate student at the University of California, worked out the necessary analytic form. By using the notation of symmetric functions he was able to give a compact representation of the formula as a triple summation, holding for a polynomial of any degree.

The method now to be presented is completely general in that it leads directly to a power series in x that exactly satisfies a series of points $f(a_0)$, $f(a_1)$, $\cdots f(a_n)$, whether equally or unequally spaced and regardless of the value of n, without the explicit use of any equation whatsoever. Furthermore, the necessary calculations can be carried out more rapidly than in the case of any alternative method. This last fact is indicated by the following considerations.

Let the addition, or subtraction, or multiplication, or division of two numbers be termed an *elementary operation*. Then if four simultaneous linear equations are solved by determinants, 311 elementary operations are in general involved. If the four equations are polynomials in x, only 235 operations are necessary. If one uses the Gauss method, in which repetitions of the same elementary operation are systematically eliminated,[4] only 62 operations are needed for four equations, in general, and 54 operations for four polynomials.

In contrast to the foregoing, the following methods employ *differences* of the original data, and the smaller number of required elementary operations is due primarily to that fact. If one has equally spaced arguments and uses the method given in reference 1,[5] 31 operations are required. If the arguments are unequally spaced, and one uses Kalbfell's equations, leading to a $f(x-a_0)$ and then by a Horner shift to the desired $f(x)$, 39 operations are required. But the method of divided differences with repeated arguments requires only 30 elementary operations for unequally spaced arguments, and only 27 operations for equally spaced arguments.

As already noted, the interpolation formula for use with a table of divided differences is known as Newton's equation. If the argument a_0 in such a table is repeated, Newton's equation takes the form of a power series in $(x-a_0)$ which, as proved by Milne-Thomson,[3] is just a Taylor's expansion. Hence the special divided differences appearing as coefficients of $(x-a_0)^i$ are merely the respective derivatives of the function at $x=a_0$ except for a factor $i!$ (see Eq. (4) ahead).

If, now, one extrapolates a table of divided differences to $x=0$, and then repeats this zero argument, it is possible to obtain directly the desired power series in x. This last point, although implicit in references 2 and 3, is not stated explicitly by either, and possibly for this reason seems to have escaped the notice of scientists.

In order to make the new method as clear as possible, I first review the meaning of divided differences.[6] One is given $n+1$ entries $f(a_0)$, $f(a_1)$, $\cdots f(a_n)$, with arguments $a_0a_1\cdots a_n$. Then a table of divided differences has the form of Table I.

The *first* divided differences are defined by

$$f(a_0a_1)=\frac{f(a_0)-f(a_1)}{a_0-a_1},\quad f(a_1a_2)=\frac{f(a_1)-f(a_2)}{a_1-a_2},\quad \text{etc.}$$

The *second* divided differences are defined by

$$f(a_0a_1a_2)=\frac{f(a_0a_1)-f(a_1a_2)}{a_0-a_2},$$

$$f(a_1a_2a_3)=\frac{f(a_1a_2)-f(a_2a_3)}{a_1-a_3},\quad \text{etc.}$$

TABLE I. Divided differences.

a_0	$f(a_0)$			
		$f(a_0a_1)$		
a_1	$f(a_1)$		$f(a_0a_1a_2)$	
		$f(a_1a_2)$		$f(a_0a_1a_2a_3)$
a_2	$f(a_2)$		$f(a_1a_2a_3)$	
		$f(a_2a_3)$		
a_3	$f(a_3)$			

[4] The standard method of solving simultaneous linear equations by the successive elimination of unknowns, if systematically carried out, is just the Gauss method.

[5] *i.e.*, a standard table of central differences, then Eq. (1), and finally the shifts from $f(\epsilon)$ to $f(e)$ to $f(x)$.

[6] See reference 2, pp. 20–28.

TABLE II. Divided differences with repeated arguments.

a_3	$f(a_3)$			
		$f(a_2a_3)$		
a_2	$f(a_2)$		$f(a_1a_2a_3)$	
		$f(a_1a_2)$		$f(a_0a_1a_2a_3)$
a_1	$f(a_1)$		$f(a_0a_1a_2)$	
		$f(a_0a_1)$		$f(a_0a_0a_1a_2)$
a_0	$f(a_0)$		$f(a_0a_0a_1)$	
		$f(a_0a_0)$		$f(a_0a_0a_0a_1)$
a_0			$f(a_0a_0a_0)$	
				$f(a_0a_0a_0a_0)$
a_0				

In general

$$f(a_0\cdots a_n)=\frac{f(a_0\cdots a_{n-1})-f(a_1\cdots a_n)}{a_0-a_n}.$$

Such a table can be used to interpolate any value of $f(x)$, and the result is given by *Newton's equation*

$$f(x)=f(a_0)+(x-a_0)f(a_0a_1)+(x-a_0)(x-a_1)f(a_0a_1a_2)+(x-a_0)\times(x-a_1)(x-a_2)f(a_0a_1a_2a_3)+\text{etc.} \quad (2)$$

The continuation of this equation is obvious.

In the extension of a table of divided differences by the use of repeated arguments, we merely repeat the value of a certain argument a specified number of times. This repeated argument may be one of the original given arguments, or it may be an interpolated or extrapolated value, including zero. Furthermore, rearranging the order of arguments in a divided difference table does not affect the final result. Hence the repeated argument can be placed at any point in the table. It is normally placed at the top or at the bottom of the table.

If $a_0a_1a_2\cdots$ represent ascending values of the argument, it is most convenient, for our present purpose, to invert the order in Table I and place a_0 at the bottom of the table. Moreover, it is customary to designate the repeated argument by a_0 although, as noted, a_0 may represent one of the given arguments, or any other desired value. A table of divided differences with repeated arguments then takes the form of Table II.

The divided differences occurring in Newton's equation (2) then form the *ascending* diagonal $f(a_0)$ to $f(a_0a_1a_2a_3)$. Those forming the *descending* diagonal $f(a_0)$ to $f(a_0a_0a_0a_0)$ are needed for the desired power series form

$$f(x)=f(a_0)+(x-a_0)f(a_0a_0)+(x-a_0)^2f(a_0a_0a_0)+(x-a_0)^3f(a_0a_0a_0a_0). \quad (3)$$

If one uses the standard definition of a divided difference, each of the divided differences in Eq. (3) has the indeterminate value 0/0. But as already noted, these quantities may be shown to be proportional to the respective derivatives, at $x=a_0$, and Eq. (3) is, in fact, just Taylor's Expansion. Thus for n coincident (i.e., repeated) entries

$$f(a_0a_0\cdots a_0)=f^{(n-1)}(a_0)/(n-1)!, \quad (4)$$

where $f^{(n-1)}$ is the $(n-1)$st derivative of $f(x)$ at the point a_0.[3]

The standard use of a table of divided differences is to obtain an interpolated or extrapolated value of the function. Since the first step in the direct use of the table to obtain the function in power series form consists in just such an operation, we first illustrate it. To obtain the desired numerical result we may make an explicit substitution in Newton's equation (2). We may also perform the necessary calculations directly in the table of differences, and it is this latter form of procedure that represents the first step in setting up the power series form.

Let us assume the following set of four entries, unequally spaced, which may obviously be satisfied by a third degree polynomial,

$x=$	3	4	6	9
$y=$	26	53	173	608

As a first problem we evaluate y for $x=2$.

For all processes discussed here, it is convenient to invert the order of arguments, placing the smallest argument at the bottom of the table. It is also convenient to reverse both the numerator and denominator, in each divided difference, in order to obtain positive quantities in the case of an ascending function. The result is Table III, in which the calculation of each divided difference is shown in detail.

Newton's equation (2) uses the backward diagonal based on a_0, consisting of the numbers 26, +27, +11, +1. Thus Eq. (2) becomes

$$f(2)=26+(2-3)(+27)+(2-3)(2-4)(+11)+(2-3)(2-4)(2-6)(+1)=+13. \quad (5)$$

The second method consists of the direct use of Table III, and the new quantities thus calculated are given in italics. The basis of every interpolation formula is the assumption that a certain order of difference is constant. Here the single known third difference is assumed constant, in order to obtain the desired extrapolated value. We thus obtain the italic $+1$ in the last column. From its definition, this quantity results from the division of "something" (indicated by a star) by $(6-2)$. Hence the "something" is $+4$, which is placed in the second difference column and subtracted algebraically from $+11$ to get $+7$. This second difference, in turn, must result from "something" divided by $(4-2)$. We thus get $+14$, which is placed in the first difference column and subtracted from $+27$. Continuation of the process yields $f(x)=f(2)=+13$.

We now use the same set of entries, but calculate the resulting third degree polynomial in power series form (Solution 1) and also in five other possible forms. This variety of possibilities results from the fact that the symbols $a_0a_1a_2$ of Newton's equation can be identified in any desired way with the given arguments. In fact, if the series of repeated arguments, all labelled a_0 in Eq. (3) and in Table II, are labelled respectively $a_0a_1a_2$ etc., then Eq. (3) becomes identical with Newton's equation (2). Hence Eq. (2) can be used for any sort of calculation, by merely identifying $a_0a_1a_2$ etc., with the appropriate numerical arguments.

To obtain the result directly in power series form, the repeated argument must be *zero*, and the first step in the process is to evaluate the function for $x=0$. We then repeat the (assumed constant) nth divided difference n additional times and use the numerical process just outlined to calculate the various needed divided differences—here $f(a_0a_0a_0)$, $f(a_0a_0)$, and $f(a_0)$ of Eq. (3). As before, all such calculated values are shown in italics, in Table IV, with the table of divided differences resulting directly from the given entries shown in Roman characters.

The desired power series form may be obtained directly from Eq. (2) by identifying the three repeated arguments with a_0, a_1, and a_2, respectively. We thus get a solution that uses the lowest diagonal and yields, from Eq. (2)—or from Table II and Eq. (3)—

$$\text{Solution (1)} \quad y=f(x)=5+4x-2x^2+x^3. \quad (6)$$

The main purpose of the paper is to outline this very rapid and convenient way of obtaining the desired Eq. (6). Examination of Table IV shows that only 18 elementary operations are performed on the Roman numerals, and 12 on the italicized numerals, giving a total of 30, as already stated. If the repeated argument has a finite value a_0 in place of zero, there are 6 additional elementary operations, leading to a

TABLE III. Evaluation of y for $x=2$.

Argument	Entry	First difference	Second difference	Third difference
$a_3=9$	608			
$(+3)$	$(+435)$	$\frac{+435}{+3}=+145$		
$a_2=6$	173		$(+85)$ $\frac{+85}{9-4}=+17$	
$(+2)$	$(+120)$	$\frac{+120}{+2}=+60$	$(+6)$	$\frac{+6}{9-3}=+1$
$a_1=4$	53		$(+33)$ $\frac{+33}{6-3}=+11$	
$(+1)$	$(+27)$	$\frac{+27}{+1}=+27$	$(+4)$	$\frac{*}{6-2}=+1$
$a_0=3$	26		$(+14)$ $\frac{*}{4-2}=+7$	
$(+1)$	$(+13)$	$\frac{*}{+1}=+13$		
$x=2$	13			

TABLE IV. Power series solution.

Solutions 1	2	3	4	5	6	x	y	First difference	Second difference	Third difference
				a_2	a_2	9	608			
						$(+3)$	$(+435)$	$\frac{+435}{+3}=+145$		
			a_2	a_2	a_1	6	173	$(+85)$	$\frac{+85}{9-4}=+17$	
						$(+2)$	$(+120)$	$\frac{+120}{+2}=+60$	$(+6)$	$\frac{+6}{9-3}=+1$
		a_0	a_2	a_1	a_0	4	53	$(+33)$	$\frac{+33}{6-3}=+11$	
						$(+1)$	$(+27)$	$\frac{+27}{+1}=+27$	$(+6)$	$\frac{*}{6-0}=+1$
	a_0	a_1	a_1	a_0		3	26	$(+20)$	$\frac{*}{4-0}=+5$	
						$(+5)$	$(+21)$	$\frac{*}{5-0}=+7$	$(+4)$	$\frac{*}{4-0}=+1$
a_0	a_1	a_2	a_0			0	+5	$(+5)$	$\frac{*}{5-0}=+1$	
								+4	$(+3)$	$\frac{*}{3-0}=+1$
a_1	a_2					0			−2	
										+1
a_2						0				

power series in $(x-a_0)$. Finally, if the arguments are equally spaced, only 27 elementary operations are needed to derive a third degree power series in x.

The remaining possible five forms of this same power series, which exactly satisfies the given entries, are obtained by direct substitution in Eq. (2), by identifying the several numerical arguments with the a_0a_1 and a_2 of that equation, in the ways indicated in Table IV. Note that a_3 occurs only in connection with the final divided difference, $f(a_0a_1a_2a_3)$.

Solution (2)

$$y=26+7(x-3)+x(x-3)+x^2(x-3). \quad (7)$$

Solution (3)

$$y=53+27(x-4)+5(x-4)(x-3)+x(x-4)(x-3). \quad (8)$$

Solution (4)

$$y=5+7x+5x(x-3)+x(x-3)(x-4). \quad (9)$$

Solution (5)

$$y=26+27(x-3)+11(x-3)(x-4)+(x-3)(x-4)(x-6). \quad (10)$$

Solution (6)

$$y=53+60(x-4)+17(x-4)(x-6)+(x-4)(x-6)(x-9). \quad (11)$$

Solution (5) is the ordinary polynomial in factorial form, as used for interpolation. It has already been evaluated in Eq. (5), for $x=2$. The other four solutions are of little practical value and are given merely to indicate the inherent possibilities of Table IV. The two important solutions are (1) and (5). The only form of solution not illustrated by Table IV is the power series in $(x-a_0)$. Such a power series is obtained by using a_0 for the repeated argument, in place of zero, and carrying out the same series of operations in Table IV.

In conclusion it may be noted that Eq. (2) is the basis of all solutions. But in the case of Solution (1), the resulting form of the equation—i.e., Eq. (3)—is so simple that there is nothing to memorize. The lowest diagonal of an extended table of divided differences, such as Table IV, gives directly the respective coefficients of the power series in x, regardless of the degree of the polynomial. This is the basis of the statement made earlier that no equations are explicitly needed for the method advocated here for setting up a power series in x. Another very real advantage of the method results from the fact that it deals always with *differences* and hence leads in general to considerably smaller numbers than does a straightforward elimination of successive unknowns (Gauss method). Thus in the sample problem just given, the 54 numbers resulting from the 54 elementary operations of the Gauss method average 108.39, whereas the 30 numbers in Table IV, resulting from the 30 elementary operations shown in that table, average only 35.60.

An entirely obvious extension of the foregoing method for setting up a power series in x is to the case of the evaluation of derivatives, based on a given set of data. One can find in the literature more or less complex expressions for the first, second, third, etc. derivatives, in terms of (a) finite leading differences, (b) finite central differences, and (c) divided differences.[7] The divided difference formulas are particularly involved, but such formulas must be used when the entries are unequally spaced. One may, however, avoid the use of any and all of these formulas by merely setting up the appropriate power series in x and differentiating it. As an example, if one has a set of five entries, a fourth degree power series in x will yield derivatives of various orders that are just as accurate, for a point close to the center of the set, as the published analytic formulas for the respective derivatives, carried through the fourth finite difference.

In the case of numerical integration, the standard methods given in the literature (Newton-Cotes formulas, and remainder type formulas) correspond just to the setting up of an appropriate power series in x and its integration between defined limits. If the entries are equally spaced it is undoubtedly simpler to make direct use of the published formulas. But when the entries are unequally spaced—a case that is normally not even mentioned—one may solve the problem quite simply by setting up a power series in x (as in Table IV) and integrating it.

[7] Reference 2, pp. 62–63, pp. 64–65, and pp. 66–68 respectively.

The Physicist and Scientific Method

WILLIAM H. GEORGE
Chelsea Polytechnic, London, S.W. 3, England

SCIENCE is both a body of knowledge and also a type of action. Scientific method deals with this latter aspect of science. Using military terms, President Conant has called it "the tactics and strategy of science." The physicist as such is not uniquely concerned with scientific method. But neither is the chemist nor is any worker in the biological sciences. For that reason alone the study of the subject is apt to be neglected by the very workers who by their typical actions as physicists or chemists or biologists practice it most assiduously. It may be that those who can, do and those who cannot, teach. But if so, do they teach scientific method?

In the analytical subject indexes of the *American Journal of Physics* provision is made for a section on scientific method. Nevertheless, in the most recent volume, No. 16 for 1948, there is not a single entry under this heading in spite of the fact that articles are indexed on analyses of the contents, rather than on the mere titles. The purpose of the present article is to direct attention to a subject highly relevant to both the training of research students and to the presentation of science to the nonspecialist student. Co-operation is also sought from other workers interested in the development of the subject. It should be distinctly understood that although this discussion is confined to physics, from the very nature of scientific method, parallel discussions apply to each of the sciences.

For clarity and brevity a series of isolated statements will be made and then we will try to fit them into a unified whole.

(1) The majority of students who pass through the physics department of a college or university do not become physicists.

(2) The reputation of a physics department depends upon its research. Yet the few exceptionally brilliant research students are the very students who are least dependent upon methods of study or teaching. They take to research as a duck takes to water. Unfortunately, these "natural" research workers who are fertile in physics are too few to satisfy the needs of present-day civilization. They must be aided by the work of many more of the less creative workers who, in their turn, are far more dependent upon training.

(3) If scientific action "comes natural" to a research worker, he may have no conscious knowledge of scientific method and may, therefore, have no interest in the subject. Striking exceptions in physics are Newton, Faraday, and Helmholtz, to mention only three who have written very clearly on scientific method.

(4) Teaching is still an art rather than a science. Nevertheless, insofar as present knowledge goes "transference of training" (or whatever phrase the reader prefers) is favored, if it occurs at all, when the student's attention is specifically directed to the thing to be transferred. That is to say, "exposure" of an average student to the subject matter of physics does not, in itself, favor the development of scientific action. Unless his attention has also been continually directed to the character of those actions which are typically scientific, his knowledge of the facts, the experimental and the mathematical techniques of physics will be no guarantee that he has conscious knowledge of scientific method.

Before adding to this set of isolated statements let us consider the first two together. Because of statements (1) and (2) it follows that the bulk of the work of a physics department is outside the work on which its reputation is based; that is, it is *not* concerned directly with the production of brilliant research physicists. Why then are all these students "exposed" to physics? One justification relates to science as a body of knowledge. They live in a world of applied science and will, we hope, use physical appliances more efficiently if they know something of the physical principles involved. Also, as responsible citizens in a democracy they will see the necessity for supporting the research upon which applications are based.

But what of the other aspect of science? Is the physics so taught that the student not only knows some physics, but that he also knows and applies the tactics and strategy of science? The answer to this question is, undoubtedly, no. After taking the average course in college physics the student is no more scientific than he was before. Such a statement does a grave injustice to the few exceptional courses which depend upon the special qualities of certain physicists well known to readers of this journal. But the majority of physics courses leave the students with a knowledge of varying quantities of physics but with no conscious knowledge of scientific method.

At this stage of the discussion addition will be made to the list of isolated statements.

(5) In the absence of any large scale experiments on the subject an open mind must be kept, but probably the most efficient method of training students to use scientific method, in action, is to teach the subject in association with one or more of the sciences or at least in association with some body of knowledge in which it is applied.

(6) The majority of physics students are given no specific instruction in scientific method. Where courses are offered at all they are given in departments of philosophy and are not normally taken by science students.

(7) Research workers are not necessarily scientific outside the laboratory. On the average, in subject matter outside the field of specialization they are neither more nor less scientific than the average citizen.

By critics of science and by some professional scientists, statement (7) is used as an argument against the cultural value of physics or of any other science. In effect they say that a science course has no other educational use but to give the student some particular technical knowledge. But statement (4), combined with statement (6), gives precisely the observed result. To teach *physics-without-scientific-method* is a policy we should use if we had no particular desire that the student should learn anything about scientific method. We fail to direct the students attention to the tactics and strategy of science and on his part he reciprocates by giving no attention to it. The wisdom of the proverb tells that you can lead a horse to water, but you cannot make him drink. Under present conditions we omit to follow even the first part of the advice.

The main conclusions so far reached are that firstly, physics teaching does not make the students scientific, and secondly, that this result *is to be expected* from the nature of the teaching and from present day knowledge of educational technique. The remedy is suggested by statement (5) which leads to the second part of our discussion. If scientific method is to be taught its character must be analyzed so that general agreement is reached on what is to be combined with the teaching of physics.

Here clarity and brevity will be helped by consideration of two more isolated statements.

(8) All the standard treatments of scientific method in all languages are based on philosophy and most of them are written by authors having no experience of scientific research.

(9) As a working basis of definition usable in action, a philosophical statement is taken to be one necessitating the use of one or more of the four words *truth*, *reality*, *existtence*, *values* (beauty, goodness, etc.), or any of their grammatical derivatives. For the past 2000 years the most powerful intellects have been working upon philosophy *as here defined* and have reached complete disagreement in all the main topics. For any one "result" reached by one eminent philosopher, the opposite "result" has been reached by some other equally eminent philosopher.

The relevance of statements (8) and (9) to the present discussion cannot be too strongly emphasized. As Faraday never tired of saying throughout his life "Facts are the foundation of science." Facts, and therefore the bases of science, are in practice based upon agreement. Philosophy, as here defined, results in disagreement. Two specific examples are particularly relevant for physicists. Firstly, when mathematical physicists of the caliber of Eddington and Jeans mix physics with philosophy their results are still heavily

criticized by professional philosophers. Secondly, the question "Can quantum-mechanical description of physical reality be considered complete?" would, on the present definition, be considered philosophical. The result of a discussion between Einstein and Bohr[1] is disagreement.

Philosophy and logic alike are critical rather than creative. Whilst they may provide wisdom *after* the event, they give little help in creativeness. The more a piece of research is creative the more we physicists like it. Now whatever else it may be, scientific method is the technique of research. Scientific method is the tactics and strategy of scientific research. If we are interested in scientific action, then we must teach students not only physics, but also scientific method. To teach scientific method we must first find out what it is. The search must be scientific. Philosophy with its wordy confusion is associated with inaction. Hence, to study scientific method through philosophy alienates the research workers mentioned in statement (3) and yields a product tending to promote contemplative inaction in these students whom it does not confuse.

At the present time the United States is particularly well adapted for such a study. The President's Scientific Research Board has called for systematic investigation of the training of scientists and the Cooperative Committee on the Teaching of Science and Mathematics of the A.A.A.S. has recommended that means should be found for older scientists "to refresh their scientific spirit, and to study new methods of science education and research in new environments."

The long range program specifies "Comprehensive investigations should be undertaken to: —(1) determine the concepts and principles of science and mathematics essential to an adequate program of general education." In Europe attention is largely focused on getting the prime necessities of life. Even in England a physicist ordering laboratory equipment may well get a request for a statement of how much of the equipment is for export. Under such conditions little effort is available for a study so far removed from the prime necessities of life, or of war. The "college" course with its vigorous experiments in teaching science to the non-scientist student provides a vast field for study unavailable in Europe where science students are trained for professional work only in science.

In conclusion it should be noted that this discussion is in no way dependent upon the writer's own particular views stated elsewhere,[2] on the nature of scientific method. If the suggestions were adopted and the reader decided to teach scientific method in association with physics, he would still have to decide what was the nature of scientific method and what features of it could be incorporated into his own treatment of physics. Furthermore, the proposals would harmonize with all the varied methods now being used to teach physics. Scientific method could be incorporated as well into a historical as into a logical treatment of physics, to mention only two extreme types. It might be used as a unifying or integrating principle in elementary survey courses. At the other end of the scale little harm would be done to the advanced graduate student who proposed to spend the rest of his life working in theoretical physics if he spent a little time studying, for example, the characteristics of theories. The advanced experimental student vigorously engaged in establishing the "existence" of the latest fundamental particle might relax for a few hours to study how some predecessor established the *existence* of say, the aether—at least to the satisfaction of his generation.

To show how specially suited is the subject of physics to the study of scientific method would need a further discussion which it is hoped to present later.

[1] A. Einstein, B. Podolsky and N. Rosen, *Physical Rev.* **47**, 777 (1935); N. Bohr, *Physical Rev.* **48**, 696 (1935).

[2] W. H. George, *The scientist in action: a scientific study of his methods* (New York, 1938). This book is too long winded for general use, but contains a summary at the end. The writer would be very grateful to hear from any readers interested in the scientific study of scientific method.

If you have had your attention directed to the novelties in thought in your own lifetime, you will have observed that almost all really new ideas have a certain aspect of foolishness when they are first produced, and almost any idea which jogs you out of your current abstractions may be better than nothing.—A. N. WHITEHEAD.

An Experimental Approach to the Paraxial Properties of Lens Systems

F. D. Cruickshank
University of Tasmania, Hobart, Tasmania

THE general paraxial theory of lens systems constitutes an essential part of any study of optics. In the normal university course it is important to provide a sound basic knowledge of paraxial optics irrespective of the level to which the further study of lens systems may be carried, while for the few whose work will lie in the field of technical optics it is very obviously a basic element. Although much has been written in connection with the teaching of this and of other sections of optics, there is still need for improved presentation. The Physical Society's *Report*[1] paid little attention to the experimental side of the problem. The writer has had occasion to consider this aspect in some detail, and the following description of an experimental study of the paraxial properties of lens systems may be found of interest to other teachers of physics.

In writing this paper it is specially in mind to illustrate a method of teaching experimental physics in the junior undergraduate laboratories which has been developed with considerable success in the University of Tasmania during the past eleven years. The method aims at the introduction of something of the atmosphere of research laboratory into the teaching laboratory, and is described in detail in an earlier paper.[2] In the study of the paraxial properties of lenses in the first year course the writer has found it profitable to introduce students to the problem by means of experimental work with a thick lens, or lens system, of positive power. Such a system is no more difficult to handle experimentally than a thin lens, and the general results are more useful. The thin lens may be considered later as an easy special case. The work is presented to the students in the form of a series of problems for investigation which develop normally one from the other.

1. Study of the Projected Images Formed by a Lens System

The property of the lens system which first attracts attention is its ability to project an inverted image of a bright object onto a piece of white card or other suitable screen. Using the lens system and a ground glass screen we can project onto the screen images of bright objects in and beyond the laboratory. For near objects it is convenient to use an optical bench, with an illuminated graticule as the object, to investigate the changes in the position and size of the image as the object is moved. We notice during these preliminary experiments that for each object position there is one corresponding screen position for which a sharply focused image will be formed. Images of a distant cloud, or the sun, can be projected on the screen by the lens. It is thus evident that there is no maximum distance of the object from the lens for which an image of it can be formed on the screen by the lens. As we move the object towards the lens the projected image recedes from the lens, increasing in size, and it is possible to continue this process until, finally, the image may be projected onto the opposite wall of the laboratory. It would seem just to infer that this trend would continue until the image was formed at a very great distance and have a very great size.

These preliminary observations suggest several matters which can be investigated quantitatively, the most obvious being (i) the relation between the positions of the object and its image relative to the lens system, (ii) the relation between the size of the image and its position relative to the lens, and (iii) the relation between the size of the image and the position of the object relative to the lens. The experimental determination of the position and size of the image for various object positions will provide observational data from which the character of these relationships may be deduced.

As a first step towards the elucidation of these relationships we plot whatever graphs we can from the experimental data obtained as it is easier to derive information from graphs than from columns of numbers. It is convenient to deal with the image size by using the ratio of the image size to the object size, which we call the magnification m' of the image. Our experi-

[1] The Physical Society, *Report on the teaching of geometrical optics* (London).
[2] F. D. Cruickshank, *Am. J. Physics* **17**, 15 (1949).

ments show that m' increases with the distance l' of the image from the back surface of the lens, and decreases with the distance l of the object from the front surface of the lens. These observed trends suggest making the plots shown in the lines marked (*a*) in Figs. 1 and 2. In Fig. 1 the graph has been extrapolated to cut the l' axis at a point for which $l' = 10.2_0$ cm which is the image distance for zero magnification. This extrapolation is clearly legitimate, for the decreasing m' values correspond to increasing values of $1/m'$ which are shown in Fig. 2 to be associated with increasing distance of the object from the lens, and thus represent a continuation of the experiment to greater object distances. Zero magnification appears as the limiting value of m' as the object becomes infinitely distant, so that the intersection point $l' = 10.2_0$ cm, thus defines the position of the image for an infinitely distant object. Similar reasoning shows that the extrapolation in Fig. 2 is legitimate and that the intersection point $l = 12.4_0$ cm specifies the object position corresponding to a projected image infinitely large compared with the object, and infinitely distant.

It is useful to summarize the preceding matters diagrammatically, and this is done in a part of Fig. 3. In this diagram I and I' are infinitely distant axial points, and F and F', usually called the first and second focal points, are the axial intersection points, $l = 12.4_0$ cm and $l' = 10.2_0$ cm. For small plane objects anywhere between I and F, which we may call the region S_1, a projected image will be formed by the lens at some position between F' and I', the region S_1'. Thus we gain the concept of two regions of space associated with the lens, an object region and an image region, the boundaries of which intercept the axes of the lens at different distances on either side of the lens, but extend to infinity along the axis away from the lens. The equations of the straight lines may now be written down, if desired, but it is easier to proceed graphically for the present and consider the algebraic relations later.

2. Study of the Virtual Images Formed by the Lens System

It is natural to inquire now whether there is any process of image formation when the object is placed nearer to the lens than the point F. When this is investigated on the optical bench we find that if the graticule is between F and the first surface of the lens (region S_2), we can see an image of it on looking towards it through the lens. That it is an *image* is quite clear from the fact that it changes in size as the position of the object is varied. We should now investigate the extent of the region, S_2', in which these virtual images are formed, and endeavour to determine the relations between the image and object positions and the magnification. The problem now arises as to how a virtual image is to be located and its size measured when its essential property is that it can be seen on looking through the lens. The logical solution

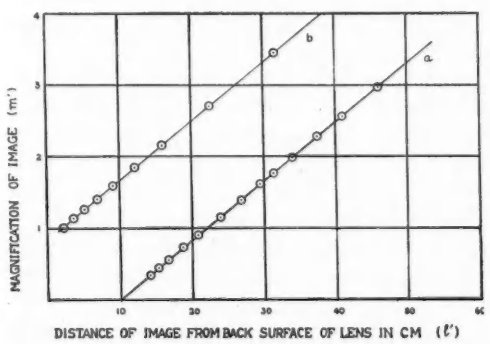


FIG. 1. Relation between the magnification of the paraxial image and its distance from the back surface of a thick lens. Line *a* for real images; line *b* for virtual images.

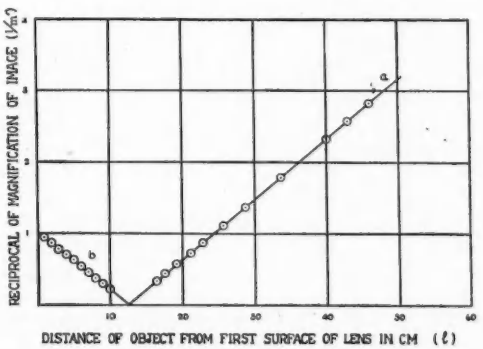


FIG. 2. Relation between the reciprocal of the magnification of the paraxial image and the distance of the real object from the first surface of a thick lens. Line *a* for real image formation; line *b* for virtual image formation.

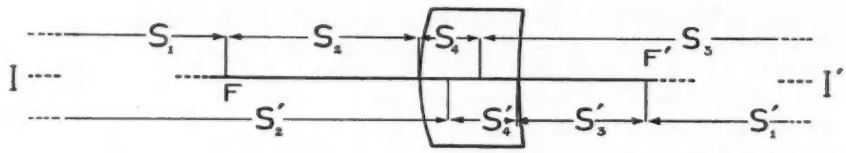


FIG. 3. Diagram showing the subdivisions of the object space and the corresponding subdivisions of the image space for the thick meniscus lens.

of this problem is described by Scott,[3] and it is an interesting indication of the value of the method of teaching being illustrated that this new and effective method of experimenting with virtual images arose directly from the investigations of a first-year class faced with this very problem.

Experimenting in this way we obtain a set of data from which the lines marked b in Figs. 1 and 2 are plotted. It is clear that the line may be extrapolated to cut both axes in Fig. 2 as these intersection points define the limits of the region S_2 which is being investigated. In Fig. 1, however, although extrapolation is legitimate indefinitely in the direction of increasing l' and m' values, it is valid only for a limited distance in the opposite direction, i.e., until the image position corresponding to the object position $l=0$ is reached. From Fig. 2 we have $l=0$, $1/m'=1.04$, whence $m'=0.96$, and the extension of the line b in Fig. 1 ends therefore at the point $m'=0.96$, $l'=1.45$. The image region S_2' thus extends along the axis in front of the lens from infinity up to a point inside the lens distant 1.45 cm from the back surface, or 0.89 cm from the front surface, the thickness of the lens being 2.34 cm. These findings are now added to the diagram of Fig. 3 by showing the extent of S_2 and S_2'.

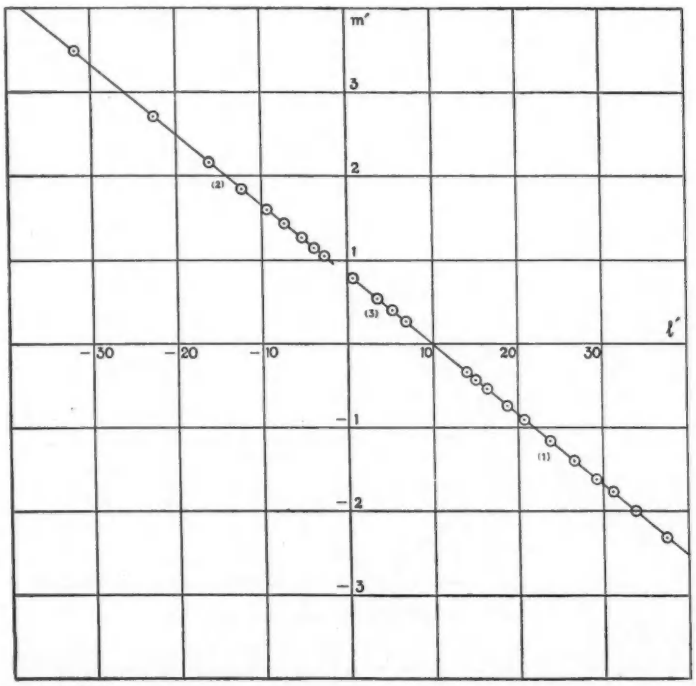


FIG. 4. Plot of the magnification m', of the paraxial image *versus* the distance in cm l', of the image from the back surface of the lens, using the sign convention. Segments (1) to (3) are from the first three stages of the investigation.

[3] B. I. H. Scott, *Am. J. Physics* **17**, 209 (1949).

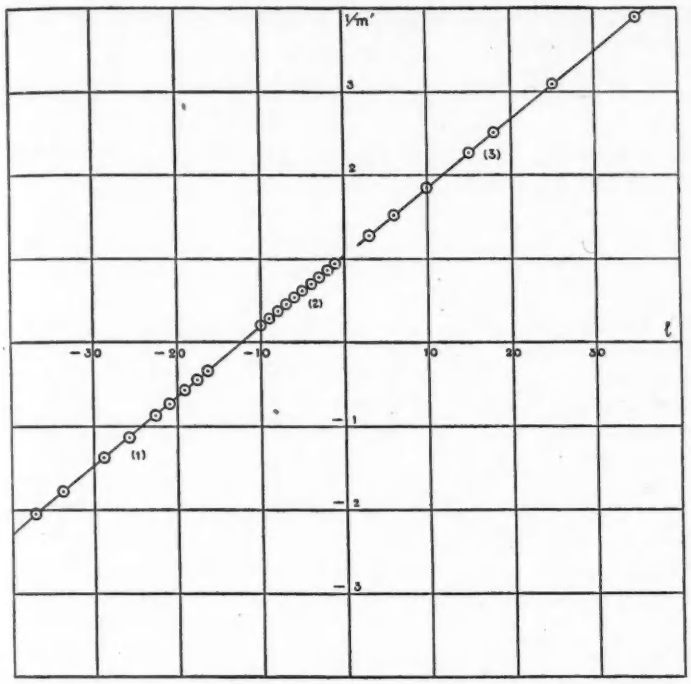


FIG. 5. Plot of the reciprocal of the magnification $1/m'$ of the paraxial image *versus* the distance in cm l of the object from the front surface of the lens, using the sign convention. Segments (1) to (3) are from the first three stages of the investigation.

3. Need for a Sign Convention

It is an aim in physical science to reduce the statement of relationships between physical quantities to the fewest number and the simplest form. The parallelism of the lines in Fig. 1 and the common intersection point in Fig. 2 suggest at once that collinearity in each of the graphs may be achieved by the adoption of a sign convention, and thus a single relation may be found to connect the pairs of variables in both sets of experimental observations. Introducing for the object specification a rectangular coordinate system with its origin at the pole of the first surface and the l axis along the optical axis of the lens, and for the image specification a similar system with its origin at the pole of the last surface of the lens, a replotting of the data of the two experiments gives the collinear segments marked (1) and (2) in Figs. 4 and 5. To be led to the adoption of a sign convention as a practical requirement emerging naturally from simple experiments is of considerable interest to the student, for this very matter has introduced much obscurity and confusion in the teaching of geometrical optics in the past, to which the amount of attention given to this point in the Physical Society's *Report*[1] bears adequate witness. It is an easy matter to introduce any particular sign convention at this stage by the appropriate twist in the argument.

4. Completion of the Space Diagram

A study of the space diagram at this stage shows that no images have been located as yet between the boundary at S_2' inside the lens and the second focal point, F'. Correspondingly there is a gap in the graph in Fig. 4, while in Fig. 5 the graph does not cross the $1/m'$ axis. These matters suggest a further investigation as to whether image formation is possible in this region. One method of proceeding is to assume tentatively that image formation is possible and that it will lead to extension of the linear graphs. The consequences of this assumption are then examined and put to experimental test. Proceeding in this way we predict from the graphs that if images are formed between the back surface of the lens and F' they should be real

and upright with a magnification less than unity, and that the corresponding object positions should be behind the front surface of the lens system. This must clearly involve the use of virtual objects. Testing this experimentally we arrange an auxiliary lens to form a real image of the graticule on the bench. If, now, we interrupt the pencils of light, which form this real image, during the path of their convergence by intercepting them on the front surface of the experimental lens, then the plane array of points to which they would have converged if allowed to do so is the virtual object for the experimental lens. By this means it is possible to produce real images in the region, S_3', between the back surface of the lens and F'. These observations add the segments marked (3) to the graphs of Figs. 4 and 5. Examining the matter of extrapolation in this case shows that the virtual object region S_3 associated with the image region S_3' extends from $l=1.75$ cm to infinity, and these regions are then added to the diagram of Fig. 3.

It is now clear from the remaining gaps in the graphs and in the space diagram that there is one further stage of investigation in which we would expect that for virtual objects inside the lens between the pole of the first surface and boundary of S_3 the lens would form virtual images between the boundary of S_2' and the pole of the last surface of the lens. The investigation of these regions calls for some very good experimenting on the part of the student, and one or two points suffice to complete the graphs, and thus the space diagram also.

5. Equations of the Graphs

Having completed the investigation of the image-forming properties of the lens we should write down the algebraic statement of the relations between the physical quantities investigated. Careful measurement shows that the slopes of the two straight lines in Figs. 4 and 5 are opposite. Denoting these by $\mp C$, and using the symbols l_f and l_f' for the distances of F and F' from the poles of the first and last surfaces of the lens, we have for the equations of the lines

$$m' = -C(l' - l_f') \tag{1}$$

$$1/m' = C(l - l_f). \tag{2}$$

Elimination of m' from these two equations gives

$$(l' - l_f')(l - l_f) = -1/C^2, \tag{3}$$

which is the equation to the hyperbola obtained by plotting l' against l. It is obvious that simple changes of variable will simplify these expressions. Thus, putting $x = l - l_f$, and $x' = l' - l_f'$, we obtain

$$m' = -Cx', \tag{4}$$

$$1/m' = Cx, \tag{5}$$

and

$$xx' = -1/C^2. \tag{6}$$

The quantities x and x' are simply the object and image distances measured along the axis of the lens from the points F and F' instead of from the poles of the surfaces of the lens.

Equations (4) to (6) show that, provided we use the focal points as origins for the measurement of the object and image distances, the image-forming properties of the lens may be described in terms of a single constant C. This must be an important quantity characteristic of the lens. Equation (4) shows that the magnification of the image depends only on the value of C and the distance of the image from the second focal point. The greater the value of C, the greater the magnification of the image produced at a *specified* distance from the second focal point. Hence, if we regard the function of the lens to be the production of magnified images of an object, it is reasonable to say that of two lenses the one which produces the image of greater magnification at a specified distance from its second focal point has the greater *power*. Thus C may be regarded as a measure of the power of the lens system. On account of the often small value of C in terms of the reciprocal centimeter, it has been found convenient to use the reciprocal of C explicitly also. We define the focal length f' of the lens as

$$f' = 1/C. \tag{7}$$

In terms of the focal length, Eqs. (4) to (6) take the more usual form

$$m' = -x'/f', \tag{8}$$

$$1/m' = x/f', \tag{9}$$

and

$$xx' = -f'^2. \tag{10}$$

6. Conclusion

It is found that once a student has made an investigation along these lines he is competent to undertake the systematic investigation of any other lens system with little or no assistance. In the writer's classes this problem is often followed by assigning to the student for unaided solution a problem such as the investigation of the object-image space relationships of a lens system of negative power. The fact that a first-year student can confidently undertake such a piece of work is evidence of the value of an experimental approach of the kind just outlined.

The problem described requires about fourteen to sixteen hours of laboratory class time. It may be considered that this is too much to devote to the properties of lens systems. If, however, we analyze the training content of the problem it will be found to be very considerable. There is a range of experimenting which begins at an easy stage and increases in difficulty as the problem advances, calling for quite some skill at the end. The adjustment and alignment of the components of the optical systems used, distance rod calibrations, etc., all involve adequate manipulative exercise. There is scope for practice in accurate measurement and repetition of observations, and it may be of interest to state that the reproducibility of the focusing operations is such that the average deviation from the mean of a series of readings is about 0.25 percent of the image distance. Thus good linear graphs are possible. The problem also affords sound training in graphical technique, the point of limited extrapolation being an instructive one. finally, there are concepts involved, such as the virtual image and object, which call for clear thinking and precise statement in the discussions which take place in the class. Our experience has shown this piece of work to be an excellent training problem in the first year. It is normally followed by a theoretical course in which the geometrical theory of lens systems for the paraxial region is developed, thus keeping the two elements of the scientific method, experimental observations and a theory developed to explain them, in their right relation. The lenses used are worked in the laboratory workshops to careful specifications, with the result that computations based on the geometrical theory confirm the experimental findings accurately, which is a source of considerable satisfaction to the student.

Experimenting with Virtual Images

BRUCE I. H. SCOTT
University of Tasmania, Hobart, Tasmania

THE difficulties involved in experimental work with virtual images are well known. It is intended to describe in this paper a method of locating and measuring the size of a virtual image which should be of interest to teachers of elementary optics. The method, apparently little known, is simple and accurate, and has been used for some time in junior classes at this university with success.

1. Description of the Method

For the present method it is to be emphasized that the virtual image produced by an optical system can be *seen* on looking through or into the system. The process of seeing involves the projection by the lens system on the retina of the eye of a *real* image of this virtual image. Hence the following procedure is suggested:—

The lens system A in Fig. 1(i) produces a virtual image I of an object scale, or graticule O. This image is used as an object for an auxiliary converging lens B, the position of which is adjusted until a clear image is formed on a ground screen S. An elementary study of lens properties will show that there is only one object position corresponding to each image position. Thus the position of I can be found very simply by removing lens A and moving O until a clearly defined image again appears on the screen. (See Fig. 1(ii).) Object scale O must now be in the position previously occupied by the virtual image I, and its location involves only one measurement.

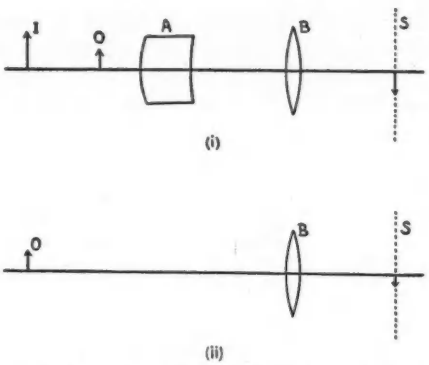


FIG. 1. Illustrating the experimental procedure in the new method. The auxiliary system (B and S) has as object (i) the virtual image I, and (ii), the object graticule O.

The magnification of the virtual image formed by the system A can be measured without difficulty. The object of the auxiliary system is first I, then O, and the ratio of the sizes of the images formed on the screen is the ratio of the sizes of I and O; i.e., the magnification of the virtual image I.

For accurate experimental work the object graticule shown in Fig. 2 has been found most suitable. It consists of concentric circles, with a central mesh of fine lines to aid focusing. This was made photographically on a lantern slide from a negative ruled on a blackened slide by means of a pantograph. The focusing screen was ground with fine emery (A.O.C. 303). A glass millimeter scale made in the same way as the graticule and held against the ground glass surface by a clip allows accurate measurements of the diameters of the rings in the image to be made.

The images produced by simple laboratory lenses usually have considerable aberration. The present method is not impeded by stopping downs the lens to an aperture at which spherical aberration is negligible, and much of its success is due to this fact. Distortion should not seriously affect the magnification measurements if the object is kept reasonably small. A considerable amount of curvature of field is tolerable provided care is taken to focus at the axial point of the image.

It is realized that this method is a variant of the more widely known telemicroscope method mentioned later in the paper. However, its greater simplicity makes it much more suitable for junior classes. It can easily be adapted for locating virtual images in mirrors. Students appear to have no difficulty in understanding or using the method if the auxiliary system is introduced to them as an *experimental eye*. It can be pointed out that the human eye cannot be used in the same way with any accuracy as the power of its lens is variable, and objects placed anywhere in a wide range of positions tend to be focused sharply on the retina.

2. Other Methods of Locating Virtual Images

At this stage it is of interest to outline other methods of locating virtual images in common use.

(*a*) *An experimental ray trace.*—From an examination of the rate of divergence of the cone of light passing through the lens due to a point source, it is possible to locate the image point in a straightforward, but rather laborious manner.

(*b*) *Methods involving parallax.*—The well-known parallax method using pins[1] is obviously too approximate for serious work. Noakes[2] de-

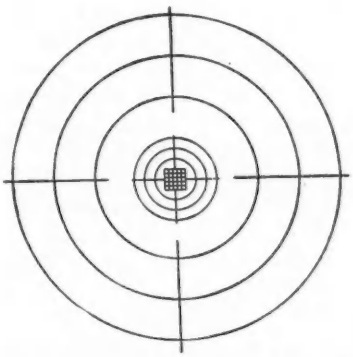


FIG. 2. A full-scale reproduction of the object graticule.

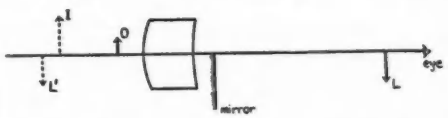


FIG. 3. The experimental arrangement in the plane mirror parallax method.

[1] Gregory and Hadley, *A class book of physics* (Macmillan), p. 292.
[2] Noakes, *A textbook of light* (Macmillan, 1937), p. 147.

scribes an improved method, the essentials of which are contained in Fig. 3. The virtual image L' of an object L in a plane mirror M is superimposed by a parallax procedure on the virtual image I to be located.

(*c*) *The telemicroscope method.*—A telemicroscope can be used for virtual image location in exactly the same way as the auxiliary system of the new method.

The accuracy of methods (*a*) and (*b*) is seriously affected if spherical aberration is present. The plane mirror parallax method will produce quite good results with care, but students take some time to master the technique and find it tiring for the eyes in a lengthy experiment. The telemicroscope method is considered unsuitable for use in the junior laboratory, where any unnecessary complication of equipment should be avoided. No significant increase in accuracy, especially in magnification measurement, is to be expected.

Results obtained by the various methods are compared with theoretical data from a paraxial trace in Table I. A thick convex meniscus lens of focal length approximately 12 cm was used. Naturally the images had considerable aberration, as it was desired to test the methods under unfavorable conditions.

Even with this lens it will be seen that very good agreement with theory can be obtained, much better than with the other methods.

TABLE I. A comparison of values of the object distance (l cm), image distance (l' cm) and magnification (m') obtained from a theoretical ray trace, and by the various experimental methods.

Object distance	Theoretical ray trace		Auxiliary system method		Plane mirror parallax method		Experimental ray trace
l (cm)	l' (cm)	m'	l' (cm)	m'	l' (cm)	m'	l' (cm)
1	2.52	1.045	2.56	1.03	3.14	1.10	2.20
2	3.69	1.144	3.68	1.13	4.34	1.18	3.40
3	5.15	1.263	5.12	1.26	5.81	1.31	4.98
4	6.91	1.411	6.84	1.41	7.48	1.52	
5	9.17	1.597	9.17	1.60	9.73	1.75	9.30
6	12.11	1.839	12.19	1.84	12.83	1.97	
7	16.10	2.171	16.15	2.17	17.15	2.36	16.50
8	21.84	2.645	22.12	2.69	24.10	3.02	
9	30.79	3.386	30.98	3.45	32.92	4.07	33.5

Moreover, the readings can be reproduced with a smaller scatter. Although the lens was stopped down, introducing a greater depth of focus, it was found that the readings at each focusing operation could be reproduced with a mean deviation of less than $\frac{1}{4}$ percent, giving the method a higher precision than the mirror method where a mean deviation of more than 1 percent was found. The readings were obtained in the shortest time and with comparatively little fatigue.

The author wishes to express his thanks to Associate Professor F. D. Cruickshank and other members of the staff of the Department for helpful criticism of this work.

RECENT MEETINGS

Western Pennsylvania Section

The regular fall meeting of the Western Pennsylvania Section of the American Association of Physics Teachers was held on December 4, 1948 at Duquesne University, Pittsburgh. Fifty members and guests were in attendance. The meeting opened with an address of welcome by Professor A. J. Kozora, Duquesne University. Professor Harry Hill, the President of the Section, was the presiding officer. The following papers were presented:

Bohr atom model. R. C. HITCHCOCK, *Westinghouse Research Laboratories.*

Network representation of input and output admittances of amplifier, LUCIO M. VALLESE, *Duquesne University.*

The Westinghouse atomic energy charts. O. BLACKWOOD, *University of Pittsburgh.*

Physics at Washington and Jefferson College in the nineteenth century. R. M. BELL, *Washington and Jefferson College.*

A carrier-current campus broadcast station. JAMES S. WILLIAMS, LESLIE COLE AND J. S. V. ALLEN, *Bethany College.*

The radian as a unit axial vector. A. G. WORTHING, *University of Pittsburgh.*

Luncheon was served in the Faculty Dining Room. At the business meeting the officers elected for 1949 were: *President,* ROBERT C. COLWELL, *West Virginia University; Secretary,* A. J. KOZORA, *Duquesne University; Vice President,* R. E. WARREN, *Thiel College.*

R. C. COLWELL
Secretary

On the Unrestricted Theorem of Carathéodory and Its Application in the Treatment of the Second Law of Thermodynamics

H. A. Buchdahl
University of Tasmania, Hobart, Tasmania

(*1*) *Introduction.* In a previous paper[1] the author has discussed Carathéodory's treatment of the Second Law of Thermodynamics from a physical viewpoint. The corresponding mathematical development makes use of the so-called *Theorem of Carathéodory*,[2] of which the author has given a proof[3] when the number of variables is restricted to three. Except in Carathéodory's own presentations[2, 4] the axiomatic treatment of the Second Law is generally carried out on the basis of the restricted theorem.[5–8] Such a procedure is, however, to some extent unsatisfactory, certainly from an aesthetic standpoint, and also because of the undesirable emphasis which it places on systems with two independent variables of state. In Part I of this paper we therefore propose to prove the Theorem of Carathéodory for the case of any number of variables by a simple extension of the author's proof of the restricted case. In Part II it is then shown how, on the basis of Carathéodory's formulation of the Second Law, the theorem provides a simple means of defining *entropy* and *absolute temperature* without the necessity of introducing either perfect gases or idealized engines and cycles.

I. Proof of Carathéodory's Theorem

(*2*) *Statement of the Theorem.* (a) In mathematical analysis of the kind under consideration it is often convenient to make use of the language of n-dimensional geometry. Thus, given a set of variables $x_1, \cdots, x_n$, we may think of a set of definite values of these, say ${}_0x_1, \cdots, {}_0x_n$, as being the coordinates of a *point*, G_0 say, in an n-dimensional picture space A into which a system of orthogonal cartesian coordinates x_i, $(i=1, \cdots, n)$, has been introduced. If $\xi_i(t)$ are continuously differentiable functions of a parameter t, then the equations

$$x_i=\xi_i(t), \quad (i=1, \cdots, n), \tag{2.1}$$

are said to specify (in parametric form) a continuous *curve* C in A which possesses a unique tangent at every point of it. If all the ξ_i's are linear functions of t, we speak of C as a *straight line*. A single relation between the x_i of the form

$$f(x_1, \cdots, x_n)=0, \tag{2.2}$$

is said to represent a *hypersurface* in A. A certain region, or part, D of A surrounding G_0 may be specified by inequalities of the kind

$$-\epsilon_i \leqslant x_i-{}_0x_i \leqslant +\epsilon_i, \quad (i=1, \cdots, n), \tag{2.3}$$

where the ϵ_i's are positive numbers. If the ϵ_i's are sufficiently small, we may speak somewhat loosely of the *neighborhood* of G_0.

All these terms are elementary extensions of the corresponding terms as used in ordinary 3-dimensional analytical geometry.

(b) The Theorem of Carathéodory may now be stated in the following form: *If in the neighborhood of any arbitrary point G_0 there are points G which are inaccessible along solution curves of the differential equation*

$$\sum_{i=1}^{n} P_i(x_1, \cdots, x_n)dx_i=0, \tag{2.4}$$

then the equation is integrable.

The functions P_i are supposed to possess continuous first partial differential coefficients with respect to the x_j's. The equation is said to be integrable if it is equivalent to a single finite algebraic relation

$$F(x_1, \cdots, x_n)=\text{const.}, \tag{2.5}$$

i.e., if there exist two functions $\lambda(x_1, \cdots, x_n)$ and $F(x_1, \cdots, x_n)$ such that

$$P_i=\lambda(\partial F/\partial x_i), \quad (i=1, \cdots, n). \tag{2.6}$$

If the solution curves to be considered are

[1] Buchdahl, *Am. J. Physics* **17**, 41 (1949). This paper will be referred to by the letter P throughout.
[2] Carathéodory, *Math. Annalen* **67**, 355 (1909).
[3] Buchdahl, *Am. J. Physics* **17**, 44 (1949).
[4] Carathéodory, *Sitz. d. Preu. Akad. d. Wiss.* p. 39 (1925).
[5] Born, *Physikalische Ztschr.* **22**, 218, 249, 282 (1921).
[6] Landé, *Handbuch der Physik* (Springer, Berlin, 1926), vol. 9, chap. 4.
[7] Chandrasekhar, *Stellar structure* (Chicago, 1939), chap. 1, p. 11.
[8] Madelung, *Die Mathematischen Hilfsmittel des Physikers* (Dover, ed. 3, 1943), p. 313.

thought of as represented in the form of Eq. (2.1), then it is assumed that the ξ_i's, but not necessarily their derivatives, are continuous functions of t, (i.e., the curves may have "kinks").

(3) *First change of variables.* In this section we carry out a change of variables in such a way that the transformation converts Eq. (2.4) into an equation in which one of the coordinate differentials dx_i does not appear. For this purpose we determine two functions $\mu(x_1, \cdots, x_n)$, $\xi(x_1, \cdots, x_n)$ such that

$$P_i=(1/\mu)(\partial\xi/\partial x_i), \quad (i=n-1, n). \tag{3.1}$$

Thus we may take ξ to be a solution of the partial differential equation

$$P_n\partial\xi/\partial x_{n-1}-P_{n-1}\partial\xi/\partial x_n=0, \tag{3.2}$$

the $x_i(i=1, \cdots, n-2)$ being regarded as constants in the process of integrating the equation, and then

$$\mu=(\partial\xi/\partial x_n)/P_n. \tag{3.3}$$

Equation (2.4) may now be written

$$\sum_{i=1}^{n-2}\mu P_i dx_i+\sum_{i=n-1}^{n}\frac{\partial\xi}{\partial x_i}dx_i=0, \tag{3.4}$$

which in virtue of

$$\sum_{i=1}^{n}\frac{\partial\xi}{\partial x_i}dx_i-d\xi=0, \tag{3.5}$$

at once gives

$$\sum_{i=1}^{n-2}(\mu P_i-\partial\xi/\partial x_i)dx_i+d\xi=0. \tag{3.6}$$

Making a change of variables from the set $x_1, \cdots, x_n$ to the set $x_1, \cdots, x_{n-1}, \xi$ we write

$$\mu P_i-\frac{\partial\xi}{\partial x_i}=Q_i(x_1, \cdots, x_{n-1}, \xi); \quad (i=1, \cdots, n-1), \tag{3.7}$$

so that Eq. (3.6) becomes

$$\sum_{i=1}^{n-2}Q_i dx_i+d\xi=0. \tag{3.8}$$

At the same time we now employ a picture space A' in which an orthogonal cartesian system of coordinates $x_1, \cdots, x_{n-1}, \xi$ has been introduced. Since we assume that within D, μ, ξ, and Q_i and their partial derivatives are single-valued and continuous functions of their arguments, it follows that there is a reversible one-to-one correspondence between the points in the range D of A and the points of the corresponding range D' of A'. Therefore, it suffices to consider Eq. (3.8) and its solutions as represented by curves in D'.

(4) *The passage along solution curves in D'.* (a) Let the initial point G_0 have the coordinates

$$({}_0x_1, {}_0x_2, \cdots, {}_0x_{n-1}, {}_0\xi). \tag{4.1}$$

(i) From G_0 pass to the point G_1 along the straight line

$$\left.\begin{aligned} x_i&={}_0x_i, \quad (i=1, \cdots, n-2)\\ x_{n-1}&={}_0x_{n-1}+({}_1x_{n-1}-{}_0x_{n-1})t\\ \xi&={}_0\xi, \end{aligned}\right\} \tag{4.11}$$

where t is a variable parameter, and ${}_1x_{n-1}$ is a constant. Accordingly $G_1(t=1)$ has the coordinates

$$({}_0x_1, {}_0x_2, \cdots, {}_0x_{n-2}, {}_1x_{n-1}, {}_0\xi), \tag{4.2}$$

where ${}_1x_{n-1}$ may be chosen at will within D'. The group of Eqs. (4.11) obviously satisfies Eq. (3.8).

(ii) Next, pass from G_1 to the point G_2 along the curve prescribed by Eq. (3.8) and the equations

$$\left.\begin{aligned} x_i&={}_0x_i+({}_2x_i-{}_0x_i)t, \quad (i=1, \cdots, n-2)\\ x_{n-1}&={}_1x_{n-1}, \end{aligned}\right\} \tag{4.3}$$

where the ${}_2x_i$ are constants. If in Eq. (3.8) the x_i's are expressed as functions of t in accordance with Eq. (4.3), the former becomes in an obvious notation

$$\sum_{i=1}^{n-2}Q_i(t; \xi)({}_2x_i-{}_0x_i)dt+d\xi=0. \tag{4.4}$$

Since this is an equation in only two variables it always possesses an integral, which we write

$$\left.\begin{aligned} \xi&=\psi(t; {}_0x_1, \cdots, {}_0x_{n-2}, {}_2x_1, \cdots, {}_2x_{n-2}, {}_1x_{n-1})\\ &=\psi(t), \text{ for short,} \end{aligned}\right\} \tag{4.41}$$

the constant of integration having been so chosen that

$${}_0\xi=\psi(0). \tag{4.42}$$

Therefore

$${}_2\xi=\psi(1), \tag{4.43}$$

and the coordinates of G_2 are

$$({}_2x_1, \cdots, {}_2x_{n-2}, {}_1x_{n-1}, {}_2\xi), \tag{4.5}$$

where the ${}_2x_i (i=1, \cdots, n-2)$ may be chosen at will within D'. In particular, ${}_2\xi$ and $\partial_2\xi/\partial_1 x_{n-1}$ are[9] continuous functions of the variable parameter ${}_1x_{n-1}$.

(iii) Finally, pass from G_2 to the point G along the straight line

$$\left.\begin{aligned} x_i &= {}_2x_i, \quad (i=1, \cdots, n-2) \\ x_{n-1} &= {}_1x_{n-1}+({}_3x_{n-1}-{}_1x_{n-1})t \\ \xi &= {}_2\xi, \end{aligned}\right\}, \tag{4.51}$$

where ${}_3x_{n-1}$ is a constant. The relations (4.51) obviously satisfy Eq. (3.8). If $t=1$ corresponds to the point G, the latter has the coordinates

$$({}_2x_1, \cdots, {}_2x_{n-2}, {}_3x_{n-1}, {}_2\xi), \tag{4.6}$$

where ${}_3x_{n-1}$ may be chosen at will within D'.

(b) Consider the equation

$$\partial_2\xi/\partial_1 x_{n-1}=0. \tag{4.7}$$

We now show that in virtue of the hypothesis contained in the theorem this equation must be satisfied *identically* in D'. For suppose that it is not identically satisfied. Then we choose G_0 (which is an *arbitrary* point!) such that

$$(\partial_2\xi/\partial_1 x_{n-1})_{{}_1x_{n-1}={}_0x_{n-1}} \neq 0. \tag{4.71}$$

In view of the assumed continuity conditions it follows from Eq. (4.43) that there exist numbers $\epsilon_i (i=0, 1, \cdots, n-2, n)$ such that for any ${}_2x_i ({}_0x_i-\epsilon_i \leqslant {}_2x_i \leqslant {}_0x_i+\epsilon_i;\ i=1, \cdots, n-2)$ we can determine ${}_1x_{n-1} ({}_0x_{n-1}-\epsilon_0 \leqslant {}_1x_{n-1} \leqslant {}_0x_{n-1}+\epsilon_0)$ such that ${}_2\xi$ takes on any prescribed value lying between the limits ${}_0\xi-\epsilon_n$ and ${}_0\xi+\epsilon_n$. Hence if, for some positive number ϵ_{n-1} we take ${}_0x_{n-1}-\epsilon_{n-1} \leqslant {}_3x_{n-1} \leqslant {}_0x_{n-1}+\epsilon_{n-1}$, we see at once that certainly all points G in the neighborhood of G_0 lying within the range

$$\left.\begin{aligned} -\epsilon_i \leqslant x_i - {}_0x_i \leqslant +\epsilon_i, \quad (i=1, \cdots, n-1) \\ -\epsilon_n \leqslant \xi - {}_0\xi \leqslant +\epsilon_n \end{aligned}\right\} \tag{4.8}$$

are accessible from G_0 along solution curves of Eq. (3.8), contrary to hypothesis. All the ϵ_i must, of course, be such that the various points considered above lie in D'. The *necessary* condition (4.7) must therefore be satisfied *identically*. Hence ψ and therefore Q_i are independent of x_{n-1}, so that the change of variables of Sec. 3 has reduced Eq. (2.4) in n variables to an equation in only $n-1$ variables of the form

$$\sum_{i=1}^{n-2} Q_i(x_1, \cdots, x_{n-2})dx_i + d\xi = 0. \tag{4.9}$$

(5) *Completion of proof by successive reduction.* (a) The procedure of the previous two sections may now be applied to Eq. (4.9). Thus we can always find two functions $\sigma(x_1, \cdots, x_{n-2}, \xi)$, $\eta(x_1, \cdots, x_{n-2}, \xi)$ such that

$$Q_{n-2}=(1/\sigma)(\partial\eta/\partial x_{n-2}), \quad 1=(1/\sigma)(\partial\eta/\partial\xi). \tag{5.1}$$

Equation (4.9) then becomes

$$\sum_{i=1}^{n-3} R_i(x_1, \cdots, x_{n-2}, \eta)dx_i + d\eta = 0, \tag{5.2}$$

with

$$R_i = \sigma Q_i - \partial\eta/\partial x_i, \tag{5.21}$$

where we have changed from the set of variables $(x_1, \cdots, x_{n-1}, \xi)$ to the set $(x_1, \cdots, x_{n-1}, \eta)$, (and to a corresponding picture space A''). By the reasoning of Sec. 4 we can then show that the R_i cannot involve x_{n-2}, so that Eq. (5.2) must be of the form

$$\sum_{i=1}^{n-3} R_i(x_1, \cdots, x_{n-3}, \eta)dx_i + d\eta = 0, \tag{5.3}$$

i.e., an equation in $n-3$ variables only.

(b) By means of altogether $n-2$ steps of the kind described above, involving $n-2$ successive changes of variables we shall therefore ultimately arrive at an equation of the form

$$Z(x_1)dx_1 + d\zeta = 0.$$

The $n-2$ changes of variables may be replaced by a single equivalent change of variables. In other words, then, from the existence of points G neighboring to an arbitrary point G_0 inaccessible from the latter along solution curves of Eq. (2.4), it follows that there exists a change of variables such that the equation reduces to the form

$$Z(x_1)dx_1 + d\zeta(x_1, \cdots, x_n) = 0. \tag{5.4}$$

But the latter has the integral (passing through G_0)

$$\int_{{}_0x_1}^{x_1} Z(x)dx + \zeta(x_1, \cdots, x_n) = \zeta({}_0x_1, \cdots, {}_0x_n). \tag{5.5}$$

[9] Kamke, *Differentialgleichungen* (Becker und Erler, ed. 2, Leipzig, 1943), p. 35.

Accordingly Eq. (2.4) is integrable, and the only points accessible from G_0 along its solution curves lie in the hypersurface represented by Eq. (5.5), which was to be proved.

The procedure above evidently affords a means of determining the integral of a linear total differential equation, if the equation is known to be integrable; and it may be noticed that it does in fact constitute a standard method, as described by Forsyth.[10]

II. Application to Thermodynamics

(6) *Direct consequences of the Principle of Carathéodory.* To apply the theorem proved in Part I we make use of that formulation of the Second Law of Thermodynamics which is known as the Principle of Carathéodory, *viz*: *In the neighborhood of any arbitrary state, J_0, of a physical system there are neighboring states J which are inaccessible from J_0 along adiabatic paths.*[11]

Let the state of some physical system of interest L be specified by a set of n independent variables

$$x_1, \cdots, x_{n-1}, t, \tag{6.1}$$

i.e., we take the empirical temperature t as one of the set.[12] The Principle of Carathéodory applies *a fortiori* to quasi-static adiabatic transitions; and these are subject to a condition of the form[13]

$$(dQ \equiv) \sum_{i=1}^{n-1} P_i dx_i + P_0 dt = 0, \tag{6.2}$$

where the P_0, P_i are certain functions of $x_1, \cdots, x_{n-1}, t$. In terms of a picture space A we may therefore say that the quasi-static adiabatics of L are the solution curves of Eq. (6.2). The principle stated above now asserts that in the neighborhood of any arbitrary point G_0 (picturing a state J_0) there are points G (states J) which are not accessible from $G_0(J_0)$ along solution curves of Eq. (6.2), i.e., along quasi-static adiabatics. Therefore, it follows at once from the Theorem of Carathéodory that Eq. (6.2) is integrable; i.e., there exist two functions λ and x_0 of the variables in the array (6.1) such that[14]

$$dQ = \lambda dx_0; \tag{6.3}$$

i.e., from the Second Law in the form above the existence of an integrating denominator of the linear differential form dQ follows immediately.

(7) *Analytical definition of entropy and absolute temperature.* (a) From the result of the previous section it follows that if L' be a second physical system, states of which are defined by a set of m independent variables, say

$$x_1', \cdots, x_{m-1}', t', \tag{7.1}$$

then in any quasi-static transition of L'

$$dQ' = \lambda' dx_0', \tag{7.11}$$

where λ' and x_0' are functions[15] of the variables (7.1).

Now let L^* be the composite system consisting of L and L' in nonadiabatic contact. Then states of L^* may be specified by the set of $n+m-1$ independent variables

$$x_1, \cdots, x_{n-1}, x_1', \cdots, x_{m-1}', t, \tag{7.2}$$

for $t \equiv t'$ when considering states of equilibrium; and such states alone are being contemplated at present. Now[16]

$$dQ^* = dQ + dQ'. \tag{7.21}$$

But, according to the considerations of Sec. 6, there must exist two functions λ^* and x_0^* of the variables (7.2) such that

$$dQ^* = \lambda^* dx_0^*. \tag{7.22}$$

Consequently, Eq. (7.21) gives rise to the

[10] Forsyth, *Differential equations* (Macmillan, London, ed. 3, 1903), p. 298.

[11] See P, Sec. 2.

[12] Let L be specified by n independent mechanical, i.e., nonthermal, variables y_i, (such as pressure, volume, electromotive force); and let L' be a second system specified by m independent mechanical variables y_i'. Then there exist two functions $t(y_1, \cdots, y_n)$ and $t'(y_1', \cdots, y_m')$ such that the two systems when brought into contact are in thermal equilibrium if, and only if, $t = t'$. This result of experience provides a definition of empirical temperature. Its scale is arbitrary since the equation $t = t'$ may be replaced by an equation $\tau(t) = \tau(t')$, where τ is an arbitrary (single-valued and monotonic) function of its argument. By selecting one particular system L and some definite function τ we then obtain a *thermometer*. To introduce the perfect gas scale at this stage is undesirable, partly because perfect gases do not occur in nature, and partly because the unique properties of this scale are themselves a consequence of the thermodynamical laws which have not yet been introduced.

[13] *Cf.* P, Sec. 5(a), and P, Sec. 3.

[14] *Cf.* P, Sec. 5(a).

[15] The λ, λ' are not uniquely determined in the sense that if λ^{-1} be an integrating factor of dQ such that $dQ = \lambda dx_0$ then clearly $[\lambda\beta(x_0)]^{-1}$ is also an integrating factor, where β is an arbitrary function of x_0. This arbitrariness is, in effect, removed later.

[16] Assuming that internal energies are additive, i.e., $U^* = U + U'$.

identity

$$\lambda^* dx_0^* = \lambda dx_0 + \lambda' dx_0'. \tag{7.3}$$

(b) It is convenient now to introduce x_0 and x_0' as two of the independent variables of state of L^*. This is legitimate since x_0 and x_0' are definite single-valued functions of the variables (6.1) and (7.1), respectively. Thus we may consider x_{n-1}, say, to be expressed as a function of $t, x_i (i=0, \cdots, n-2)$; and similarly x_{m-1}' as a function of $t, x_i' (i=0, \cdots, m-2)$. Then Eq. (7.3) becomes

$$\lambda(x_0, \cdots, x_{n-2}, t)dx_0 + \lambda'(x_0', \cdots, x_{m-2}', t)dx_0' = \lambda^* dx_0^*, \tag{7.31}$$

where λ^* and x_0^* are now functions of $x_0, \cdots, x_{n-2}, x_0', \cdots, x_{m-2}', t$. Writing Eq. (7.31) in the form

$$dx_0^* = (\lambda/\lambda^*)dx_0 + (\lambda'/\lambda^*)dx_0', \tag{7.4}$$

it is obvious that x_0^* can be at most a function of x_0 and x_0', and the same is therefore true of (λ/λ^*) and (λ'/λ^*). Since, however, λ does not involve the $x_i' (i=0, \cdots, m-2)$, nor λ' the $x_i (i=0, \cdots, n-2)$, it follows at once that λ and λ' can be functions only of x_0, t and x_0', t, respectively. The independence of (λ/λ^*) and (λ'/λ^*) of t may be expressed by the equations

$$(\partial/\partial t)(\lambda/\lambda^*) = 0, \quad \text{and} \quad (\partial/\partial t)(\lambda'/\lambda^*) = 0,$$

or

$$\partial \log\lambda/\partial t = \partial \log\lambda'/\partial t = \partial \log\lambda^*/\partial t. \tag{7.5}$$

Now the first member of this equation is independent of x_0' and the second is independent of x_0. Hence their permanent equality implies that both are the same function of t alone, $g(t)$ say. Integration of Eqs. (7.5) therefore gives

$$\lambda = e^{\int g(t)dt}\Lambda(x_0), \quad \lambda' = e^{\int g(t)dt}\Lambda'(x_0'), \quad \lambda^* = e^{\int g(t)dt}\Lambda^*(x_0, x_0'), \tag{7.6}$$

where $\Lambda, \Lambda', \Lambda^*$ are arbitrary functions[17] of their arguments. We write

$$T(t) = Ce^{\int g(t)dt}, \tag{7.61}$$

and

$$\left.\begin{aligned} S(x_0) &= \frac{1}{C}\int \Lambda(x_0)dx_0 + \text{const.} \\ S'(x_0) &= \frac{1}{C}\int \Lambda'(x_0')dx_0' + \text{const.}, \end{aligned}\right\} \tag{7.62}$$

where the constant (scale-factor) C is introduced for convenience. Accordingly, Eqs. (6.3) and (7.11) now read

$$dQ = TdS, \quad \text{and} \quad dQ' = TdS'. \tag{7.7}$$

We thus see that the linear differential form dQ which characterizes the rate at which heat enters a thermodynamic system in a quasi-static transition can be expressed as the product of a universal[18] function $T(t)$ of the empirical temperature alone and the total differential of a function S of the variables of state. T is called the *absolute temperature* and S the *entropy* of the system.

Inserting the values (7.6) and (7.7) in Eq. (7.3) the latter becomes

$$\Lambda^*(x_0, x_0')dx_0^*(x_0, x_0') = dS + dS'. \tag{7.8}$$

Using S, S' in place of x_0, x_0' as variables of state this gives

$$\Lambda^*(S, S')dx_0^*(S, S') = dS + dS'. \tag{7.81}$$

Therefore, at once,

$$\partial x_0^*/\partial S = \partial x_0^*/\partial S', \tag{7.82}$$

which shows that S and S' occur in x_0^*, and therefore in Λ^*, only in the combination $(S+S')$. Hence we define

$$S^* = \frac{1}{C}\int \Lambda^* dx_0^* \tag{7.83}$$

as the entropy of the composite system L, so that

$$dS^* = dS + dS', \tag{7.9}$$

which shows that entropy changes are additive.[19]

(8) *Irreversible processes and the Principle of Increase of Entropy.* It will be noticed that hitherto only quasi-static (i.e., reversible) processes have been considered; in particular, we were able to define entropy on the basis of these alone. Moreover, quasi-static adiabatic processes are, by Eq. (7.7), characterized by the constancy of the entropy.

To make use of the *whole* content of Carathéodory's Principle we again consider a system

[17] Arbitrary, that is to say, in the mathematical sense. They are definite functions, of course, in the case of any particular thermodynamic system.

[18] Universal in the sense that (using some definite thermometric scale t) the integrating denominator $T(t)$ of dQ is the same function of t for all thermodynamical systems, i.e., independent of the specific properties of the systems.

[19] Ultimately this is a consequence of the assumed additivity of internal energy.

L with n independent variables of state, say

$$x_1, \cdots, x_{n-1}, S, \tag{8.1}$$

i.e., we take S to be one of the variables. Let J_0 be some initial state of equilibrium of L, specified by[20] the particular values ${}_0x_1, \cdots, {}_0x_{n-1}, {}_0S$ of the set of variables (8.1). Similarly let some final state J_1 of L be specified by the values ${}_1x_1, \cdots, {}_1x_{n-1}, {}_1S$ of the variables (8.1). We inquire into the possibility of the *adiabatic* transition[21] from J_0 to J_1. Now of n variables of state we can always adjust[22] $n-1$ from one set of prescribed values to another by means of a quasi-static adiabatic process, the nth variable taking care of the constraint imposed upon the system in the form of the condition that it be adiabatically isolated. Accordingly, by means of quasi-static adiabatic operations change the variables $x_1, \cdots, x_{n-1}$ from their initial to their prescribed final values. During this process the last variable S, however, obviously remains constant. Hence we have reached the intermediate state ${}_1x_1, \cdots, {}_1x_{n-1}, {}_0S$. It remains to change S to its final prescribed value. Now if it were possible by means of any process whatever, quasi-static or non-static, to make this change in every case (i.e., irrespective of whether ${}_0S$ is greater than, equal to, or less than ${}_1S$), then obviously every state J_1 in the neighborhood of J_0 would be accessible by means of an adiabatic process. But this contradicts the Principle of Carathéodory. Hence *in any adiabatic transition of a thermodynamic system the entropy either never decreases or never increases.* By varying the initial state J_0 it is not difficult to see that the impossibility of an increase or decrease of entropy must always hold in the same sense (direction) for different initial states of a given system. The same is true for two different systems, if the additivity of entropy, as expressed by Eq. (7.9), be taken into account.

Whether the entropy can only increase or only decrease depends on the sign of the constant C in Eqs. (7.62). This is chosen so that the absolute temperature is positive.[23] One single nonstatic process then suffices once for all to decide between the alternatives. Experiment shows that in the adiabatic nonstatic expansion of gases under ordinary conditions[24] the entropy increases. Hence we have the universally valid Principle of Increase of Entropy: *In any adiabatic transition of a physical system the entropy never decreases.*

(*9*) *Conclusion.* It will have been noticed that in Carathéodory's treatment thermodynamics is built up as a kind of extension of mechanics; and in a sense the determination of the various thermodynamical functions is accordingly based explicitly on mechanical, rather than on calorimetric measurements.[25] It may be instructive to deal with a very simple example to illustrate both this point and, also, the whole preceding analysis.

Let L be a gas contained in some envelope.[26] Its state is described by the independent variables p (pressure), and v (volume), both of which are directly measurable by means of ordinary mechanical operations. Let L' be some other system with which L is temporarily in nonadiabatic contact. Let p and v be altered quasi-statically in such a way that the state of L' remains permanently unchanged. If corresponding values of p and v are plotted in a plane cartesian coordinate system, then the change of L as a whole will be represented as a continuous curve. Starting from different initial states a family of curves will be obtained densely covering the p, v plane, or at any rate a finite part of it. This family of curves may be represented by an equation of the form

$$\tau(p, v) = t, \tag{9.1}$$

every fixed value of t so giving one of the *isothermals* of L, t itself being the empirical temperature. Its scale is obviously arbitrary since the same family of curves would be represented by the equation

$$\tau(p, v) = \theta(t), \tag{9.11}$$

[20] Strictly speaking, the thermodynamic state J_0 *is* the set of constants ${}_0x_1, \cdots, {}_0x_{n-1}, {}_0S$.

[21] *Cf.* P, Sec. 4.

[22] The purely mechanical possibility of the transition is of course presupposed.

[23] The actual value of this constant is fixed by the requirement that the values of $T(t)$ corresponding to two certain states of a particular system shall differ by a prescribed amount; e.g., the system might be pure water at normal pressure, with the condition T(boiling water) $-T$(water in equilibrium with ice) $=100$.

[24] As in the Joule experiment in which a gas flows into an evacuated chamber.

[25] Implicitly this is, of course, done however we may treat the subject. The different modes of expression are simply representative of different points of view.

[26] The system consists merely of the gas. The envelope is not considered to be a part of it.

where θ is an arbitrary monotonic single-valued function of its argument.

Similarly we may measure the mechanical work done in passing by means of adiabatic, non-static processes between different states of L; and this gives rise directly[27] to the internal energy function

$$U = U(p, v). \tag{9.2}$$

On the basis of Eqs. (9.1) and (9.2), which are thus arrived at empirically, the entropy and absolute temperature functions may be determined. It is perhaps more convenient now to use in place of Eq. (9.2) the relation obtained from it on replacing the variable p by t, by means of Eq. (9.1), giving

$$U = \eta(v, t). \tag{9.3}$$

In practice one finds that in the case of common gases under ordinary conditions the equations[28]

$$\left.\begin{aligned} t &= \tau(p, v) = f(pv) \\ U(p, v) &= \text{const. } (pv) \end{aligned}\right\} \tag{9.4}$$

hold with considerable accuracy. A gas which obeys these equations *precisely* may be called an ideal gas, though such a substance is not actually found to occur in nature. The equations describing it may be written in the alternative form

$$\left.\begin{aligned} &pv = \sigma(t) \\ &U = \text{const. } \sigma(t) = c\sigma(t), \text{ say,} \end{aligned}\right\} \tag{9.5}$$

where σ is the function inverse to f, the latter depending on the selected empirical temperature scale. We now have the equation of the quasi-static adiabatics

$$dQ = dU + pdv = c\sigma'(t)dt + (\sigma(t)/v)dv = 0. \tag{9.6}$$

The latter equation possesses the integrating denominator

$$\lambda = \sigma(t)\psi(\xi), \tag{9.61}$$

where $\xi = v[\sigma(t)]^c$, and ψ is an arbitrary function[29] of its argument. Thus Eq. (9.6) becomes

$$dQ = \sigma(t)\psi(\xi)dx_0, \tag{9.7}$$

where

$$x_0 = \int d\xi/\xi\psi(\xi). \tag{9.71}$$

Now by Eqs. (7.5) and (9.61)

$$g(t) = \partial \log\lambda/\partial t = d \log\sigma(t)/dt.^{30} \tag{9.72}$$

Therefore

$$T(t) = C \exp \int (d \log\sigma(t)/dt)dt = C\sigma(t). \tag{9.8}$$

Hence, from Eq. (9.5)

$$\left.\begin{aligned} T &= Cpv, \\ U &= (c/C)T, \end{aligned}\right\} \tag{9.9}$$

so that for a perfect gas the product pv is a direct measure of the absolute temperature. Also, from the relations (7.62), (9.61), and (9.8)

$$\begin{aligned} dS &= (1/C)\psi(\xi)dx_0 \\ &= (1/C)(d\xi/\xi). \end{aligned}$$

Hence

$$S = (\log\xi)/C + \text{const.},$$

or

$$S = (\log v/C) + (c \log T/C) + \text{const.} \tag{9.91}$$

In this particular case, therefore, if v refers to the volume of one gram-molecule of the gas, then C is the reciprocal of the universal gas constant R, while c is the gram-molecular specific heat at constant volume divided by R.

Analogous considerations hold for systems with $n > 2$ independent variables of state; and when nonadiabatic processes have to be dealt with it is convenient as usual to introduce thermodynamic potentials such as the Helmholtz function F and the Gibbs function G,

$$\left.\begin{aligned} F &= U - TS \\ G &= U - TS + \sum_{i=1}^{n-1} p_i x_i, \end{aligned}\right\} \tag{9.10}$$

where the x_i are mechanical (i.e., nonthermal) independent variables of state, and the p_i are *generalized forces* such that the mechanical work done by a system in an infinitesimal quasi-static transition is given by

$$dW = \sum_{i=1}^{n-1} p_i dx_i.$$

[27] *Cf.* P, Sec. 3.

[28] The first of these is simply an expression of Boyle's law.

[29] So far we have made use only of the First Law, and there is no sound reason for making at this stage the arbitrary choice $\psi(\xi) = \text{const.}$

[30] In the differentiation x_0 and therefore ξ is to be kept constant.

Reproductions of Prints, Drawings and Paintings of Interest in the History of Physics

42. *Vanity Fair* Caricatures of HERBERT SPENCER and JOHN STUART MILL

E. C. WATSON
California Institute of Technology, Pasadena 4, California

AGAIN the question may be asked why caricatures of HERBERT SPENCER (1820–1903) and JOHN STUART MILL (1806–1873) should appear in this series of historical reproductions. The answer is twofold, (1) because of the great influence that these men exerted upon contemporary English thought and thus upon science in general and physics in particular, and (2) because it is interesting to contrast these caricatures, and particularly the written accounts that accompanied them, with those of the contemporary scientists already reproduced in this series. This contrast emphasizes the respect accorded to science and scientists during the Victorian period.

The following written statements accompanied the caricatures which originally appeared in *Vanity Fair* for 1879 and 1873:

Mr. Herbert Spencer

"HERBERT SPENCER holds the present greatest name among the philosophers. He is scarcely known in his own country outside the circles of fogies, but abroad he enjoys a wonderful reputation as the leader of all modern thought. He was born nine-and-fifty years ago, the son of a schoolmaster, who strove to culture him with the classics, and to make of him a civil engineer. But young Spencer resisted Greek and Latin, and soon renounced engineering. He was, and knew himself to be, powerful with the pen; so he became a writer. He learnt to handle a good plain English of the frigid sort. In a casual way he published articles on Government, Education, and other dull subjects, from the time he was one-and-twenty. But when Darwin invented Evolution, Evolution invented Herbert Spencer, who saw how the notion might be applied to psychologic problems. So he now addressed himself to pure philosophy, and began to publish distressing tomes. He fell foul of Comte and of Mill, and plunged about with atoms and monads in such fashion as made all men see that he must have a philosophy of his own. As nobody could well understand him his reputation waxed mightily. He is now the one recognised authority on 'Sociology;' he has discovered that 'ultimate scientific ideas are all representatives of realities

PLATE 1. "Philosophy"
[From *Vanity Fair*, April 26, 1879.]

PLATE 2. "A Feminine Philosopher"
[From *Vanity Fair*, March 29, 1873.]

that cannot be comprehended,' and that the man of science 'knows that in its ultimate essence nothing can be known.' Yet he goes on writing.

"Mr. Herbert Spencer is believed by many to be a companionable, cheerful man. He has been more than once to a shareholders' meeting to war with railway directors; he delights also in children; and he holds that suicides should rather be encouraged. Yet he goes on living."

Mr. John Stuart Mill

"In these times when men believe only what they see—even if they believe that—the attempt to deal with the evidence of things not seen is commonly regarded as an impertinent intrusion into the land of Dreams. Thought indeed does occasionally vindicate its power in strange and powerful ways to the great alarm and disquiet of those who cannot think, but it is therefore only the more feared and not the less despised. We English pride ourselves on being a practical people, who, having somehow got together certain social machines, are bent upon merely working them, and have therefore a noble impatience for the unholy audacity of a man who would ask whether better machines altogether might not be contrived. The Thinker is revolutionary in his very nature. He preaches things unheard of before—things which require attention in order to be understood, and which many of us can never understand whatever attention we give to them. He is to be dismissed therefore as a trifler, at the most to be languidly admired as a performer of lofty tumbling having no relation whatever to earthly affairs.

"It is at once the merit and the misfortune of the Thinker that he cannot be judged by his contemporaries. The Present is not his, but the Future shall be. Silently, slowly, and surely the forces he has found, if it be so, work through the generations, gathering power as they go till at last the face of the World is changed by them, and then, as it may happen, the Thinker is blessed, cursed, or ignored. John Stuart Mill has had it given to him however to influence in no small degree even his contemporary men. But that is mainly because and only in so far as he has taken and put into a more popular form the thoughts of other and preceding philosophers. He has procured the acceptation of other men's ideas, but for those which are purely his own and therefore entirely novel he has been able to do nothing beyond merely presenting them to a mocking audience. Moreover he is best fitted to be the apostle of an old rather than the originator of a new faith. No one of those who would claim to be thinkers in the less interesting branches of human affairs, save the picturesque and epigrammatic Proudhon, has written so graphically and simply, or with so much happiness (though not always correctness) of language and fertility of illustration as Mill; and that is one reason why his writings have an immediate effect. Like all thinkers he is a thorough heretic in all his beliefs —social, moral, and religious. He represents the very extreme outposts to which the most adventurous of modern thinkers have pushed, and has often found himself out of reach of the support

even of the boldest forces of Democracy. For he is a feminine philosopher, a man of vast intellect and tender feelings. He adored his wife, and since her death has lived chiefly at Avignon, in that solitude which he holds to be essential to thought. He has been an ardent champion of the principle of free exchange left to regulate itself by a free play of opposing forces; yet he makes many reservations, and only recently he has shocked his disciples by suggesting that Free Trade is an English craze not credited in other countries. He holds that there should be a 'perfect equality' of the sexes, and believes that the existing inequality between them 'must necessarily disappear,' and he supports this by the hazardous assertion that 'women—and not a few merely, but many women—have proved themselves capable of everything, perhaps without a single exception, that is done by men, and of doing it successfully and creditably.' He believes it essential to the well-being of modern communities that a 'conscientious and prudential restraint' should be placed upon the increase of population. He objects to the Ballot. He holds that minorities ought to be represented in a system of choice of legislators by majorities, and has supported schemes for new-fangled modes of boiling down the people of England into a Parliament. He declares that mechanical inventions have not lightened the toil of any human being, but have only enabled the few to become miserably richer while the many have become miserably poorer, and that 'the best state for human nature is that in which while no one is poor no one desires to be richer.' He holds that as the possessors of the 'common inheritance,' the soil of a country, are the only citizens who grow rich while they sleep, the increase of value in their lands which has arisen independently of them should be appropriated by the State for the common benefit. With these 'uncustomary opinions' he naturally regards the Tory Party as the 'stupid party' of the country, but he has nevertheless never been accepted by the Liberals as a safe exponent of their very customary opinions, and he is—as he cannot fail to be and to remain—in a position personal to himself and quite exceptional in regard to all public questions. He was born in 1806, passed his first youth as a clerk in the East India House, refused a seat on the Indian Council when the Imperial Government was substituted for that of the Company, and was elected as Member of Parliament for Westminster in 1865. He was however in an atmosphere far above that of the retired traders and society-mongers of the House of Commons, and did not even interest but only bored them; so that there was no great lamentation when in the 1868 election he was beaten by Mr. Smith, the newspaper-vendor. His influence however is greater in writing than in speech, for he is no orator and merely recites passages in a weak voice and an even tone from his books. These books will make for him an abiding reputation as one of the few men who have given themselves to thought in an unthinking generation."

New Members of the Association

The following persons have been made members or junior members (J) of the American Association of Physics Teachers since the publication of the preceding list [Am. J. Physics **17**, 165 (1949)].

Adamek, Stephen (J), 514 St. George St., London, Ontario, Canada
Alford, W. P. (J), 83 Rogers Ave., London, Ontario, Canada
Anderson, John Howard, Talladega College, Talladega, Ala.
Ashley, Ramon Lloyd (J), 2157 Northampton St., Holyoke, Mass.
Averbach, Isidor, 511 E. 48th St., Brooklyn 3, N. Y.
Bagley, Edward B. (J), 316 Bright St., Sarnia, Ontario, Canada
Barden, Elihue, Alcorn College, Alcorn, Miss.
Berntson, Charles Martin (J), 45 Fletcher St., Winchester, Mass.
Bricker, Richard McCurdy (J), 5305 Fieldcrest Dr., Pittsburgh 27, Pa.
Cake, Robert F. (J), 803 Queens Ave., London, Ontario, Canada
Carlisle, Richard T., 9 Newlyn St., Salinas, California
Carney, E. Seaton, 122 Chestnut St., Oneonta, N. Y.
Chen, Shih Chang, Physics Dept., University of Pennsylvania, Phila., Pa.
Correll, Malcolm, Gunderson 201, Oklahoma A & M College, Stillwater, Okla.
Dahler, C. E., 1906 Market St., LaCrosse, Wis.
Davis, William C. (J), 108 Powder House Blvd., Somerville 44, Mass.
Dingle, Robert L. (J), 2220 MacGregor Way, Houston 4, Texas
Discher, Wendell (J), 1443 Forbes St., Kalamazoo, Michigan
Dubs, Charles W., Box 344 Morehouse College, Atlanta, Ga.
Dudley, Howard N. III, 80 Stuntz St., Rochester 9, N. Y.
Eisele, Rev. Louis John, S.J., Spring Hill College, Spring Hill, Ala.
Eisner, Leonard, 1139 S. Atherton St., State College, Pa.
Elliott, Steward E. (J), Beta Theta Pi, Crawfordsville, Ind.
Ford, Kenneth W. (J), Palmer Physical Lab., Princeton University, Princeton, N. J.
Gardner, Edward E. (J), 7 Clinton St., Cambridge, Mass.
Garfunkel, James H. (J), 402 15th Ave. N., St. Cloud, Minn.
Gerard, Brother F. S. C., 6501 Clayton Road, St. Louis 17, Mo.
Green, Donald W. (J), 19 Williams St., Coldwater, Mich.
Hahn, Thomas M., Lt. Col., Hq. Caribbean Command, Quarry Heights, Canal Zone

(*continued on page 238*)

NOTES AND DISCUSSION

Demonstration of Lissajous' Figures

P. A. B. MEDGYESSY
Debrecen, Hungary

WITH the apparatus described below Lissajous' figures and their alteration by continuous phase displacement can be demonstrated on a projection screen. The device contains rotating parts only and is very simple in design and construction. The general idea is as follows: The curve,

$$r=k+a\cos\phi,$$

where k and a are constants, is drawn in polar coordinates on a disk. The angle ϕ is measured from any arbitrary, fixed straight line that contains the origin. As the disk is turned uniformly, the point of intersection of the curve and the straight line oscillates with simple harmonic motion along the fixed line about the point $r=k$. If a narrow strip is cut from the disk along the curve, and the disk is rotated in front of a card having another narrow straight slit, then the common opening of the slits can be projected upon a screen.

Within the limits of accuracy demanded by demonstration apparatus, a curved slit whose radius is large compared to a may be substituted for the straight slit. Another disk is therefore prepared from which a narrow slit is cut along a curve $r=k+b\cos\phi$, where b is constant. The two disks are now mounted so that the two circles of radius $r=k$ cross one another approximately at right angles, as shown in Fig. 1. It is easy to illuminate and project on a screen the point of intersection P. When one disk rotates with uniform angular velocity, and the other is stationary, the point P moves with very nearly simple harmonic motion. If both disks rotate, the point P describes the resultant of the two perpendicular harmonic motions.

In practice, the disks are mounted on axles A, B, and are coupled by gears that keep a constant frequency ratio. Excentric circles can be used instead of the correct curves without obvious loss of accuracy in the motion of the projected image.

To illustrate the effect of altering the phase of one of the components, identical pulleys are mounted on A and B and connected by an endless inextensible belt of string.

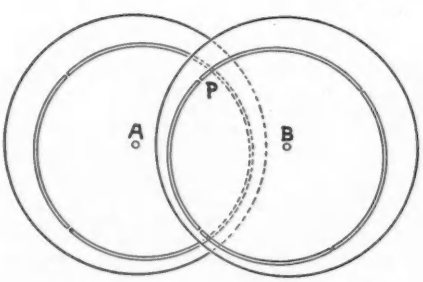


FIG. 1. Appearance of the disks used to demonstrate Lissajous' figures.

By adjusting the tension of the belt and the inertia of the driven disk, the string can be made to slip so that the disk picks up speed slowly and continuously, suffering with respect to its neighbor gradual phase displacement that can easily be observed on the screen. The change of sense in which the bright tracer spot describes the figures is easily observable.

Two methods are available for showing the effect of altering the relative frequencies of the components: Either (a) the gears by which the two shafts A and B are coupled can be changed; or (b) different disks may be used carrying slits that follow the curves $r=k+a\cos n\phi$, and $r=k+b\cos n'\phi$, where n and n' are integers, not necessarily equal.

In the present form of the apparatus, the disks were made of cardboard with $k=5$ cm, $a=0.75$ cm, and $b=0.5$ cm. The other parts were built with a "Märklin-Metallbaukasten," from which the gears were also taken.

Use of a Cathode-Ray Oscilloscope in Hoag's e/m Experiment

L. F. CONNELL, JR.
University of Texas, Austin, Texas

THE difficulties encountered by Professors McCombs and Pietenpol[1] in obtaining equipment for an experimental determination of e/m by the Busch[2] helical method have arisen at this institution also and a somewhat different method used to overcome them. A brief description of this method was submitted in November, 1948, as a note to the Editor, who then acquainted the writer with the forthcoming article by Professors McCombs and Pietenpol and suggested that the note be revised to include a discussion of the differences in the two methods.

In the original Busch method, as used by Professors McCombs and Pietenpol, the natural divergence of the electron beam produces the transverse component of electron velocity which interacts with the axial magnetic field of the solenoid and results in magnetic focusing of the electron beam. A modification of this method, introduced by Hoag[3] and described completely in his book, reduces the natural divergence of the beam to a minimum and uses an alternating potential applied across a pair of deflection plates to produce larger, but one-dimensional, transverse velocities. The cathode-ray tube screen shows a line which shortens and rotates as the magnetic field is increased and becomes a spot for exact focusing. In the Busch method the screen shows a circular spot whose size is a minimum for proper focusing. One of the principal advantages of the Hoag modification is the ease with which the focus condition may be determined.

Although it is true that the special uncoated, nonmagnetic, electrostatic-deflection tube described by Hoag has not been available for several years, satisfactory results may be obtained by use of tubes now available, such as 3AP1 and 3GP1. The deflecting plates and main elements

of almost all cathode-ray tubes of current manufacture are nonmagnetic, so that special construction for this purpose is no longer necessary. The opaque internal coating which obscures the deflection plates in most cathode-ray tubes prevents direct measurement of the distance from the deflection plates to the screen, but a method successfully used here for some time has been to determine these distances with the aid of x-ray shadowgraphs of the tube. A shadow picture may be made using x-rays from a point source several feet away and a film placed immediatley behind the small portion of the cathode-ray tube. The picture establishes the distances of the deflecting plates from some reference point on the tube base or on a visible portion of the electron gun, and the distance from this reference point to the fluorescent screen can be measured visually.

It appears pertinent to note that the 3HP7 tube used by Professors McCombs and Pietenpol was a special war-time tube which is apparently no longer available commercially.

The solution reached at this institution to the problem of additional e/m apparatus has been the use of commercial 3-inch cathode-ray oscilloscopes in the Hoag method. No modification of the oscilloscope circuits has been found necessary; hence the oscilloscopes that are in general laboratory use may be utilized. The only change involved is the removal of the cathode-ray tube from its socket in the oscilloscope and the plugging-in of an extension cord which will allow the tube to be operated inside a solenoid. The extension can easily be made from well-insulated wire and a plug and socket of the same type as the tube base. It has not been found necessary to use shielded wire. The controls on the oscilloscope provide for adjustment of brilliance, focusing, and centering and provide contacts for application of deflecting voltage to the desired set of plates. The deflection voltage used here is 60 cycle a.c. taken from the secondary of a filament transformer; the amplitude of the deflection is adjusted by the controls on the deflection amplifiers of the oscilloscope.

It should be noted that most oscilloscopes do not provide an easy means of adjusting the accelerating potential between the cathode and the second anode. A rather wide range of accelerating voltages may be obtained, however, by using an auto-transformer of the Variac type on the input to the oscilloscope and varying the input voltage within reasonable limits, say 90–130 volts, within which range most oscilloscopes will operate. The amount of the accelerating voltage, which must be known, can be measured on a high resistance voltmeter connected across the proper leads of the extension cord. If a wider range of accelerating potentials is desired than results from the above method, an external power supply can be used to raise the second anode potential at the tube.

An addition to either the Busch method or the Hoag modification, both of which call for determination of the critical field strength which will cause the electron beam to describe exactly one helical revolution, is to determine the field strength necessary to cause the beam to make exactly two, three, or any integral number of helical revolutions. The maximum field strength obtainable limits the number which may be attained. If n is the number of complete revolutions made by the electron beam, the equation for e/m becomes, in emu,

$$e/m = (8\pi^2 Vn^2/H^2L^2)\times 10^8$$

where V is in volts, H in oersteds, and L (the length of helical path) in cm.

Results obtained by use of an oscilloscope in the manner described have compared very favorably with those obtained using a set of the original Hoag-designed equipment which was available commercially[4] before the war. Averages of the thirty or more readings made by each student are usually within one percent of the accepted value.

[1] K. McCombs and W. Pietenpol, *Am. J. Physics* **17**, 78 (1949).
[2] H. Busch, *Physikalische Zeits.* **23**, 438–441 (1922).
[3] J. B. Hoag, *Electron and nuclear physics* (D. van Nostrand, ed. 2 or ed. 3), pp. 32–35 and 38–41.
[4] W. M. Welch Mfg. Co. Catalog S, Item No. 627. (1948 Catalog, but item not available.)

A Microwave Analog of a Half-Wave Plate

GENE T. PELSOR*
American University, Washington, D. C.

RECTANGULAR wave guides are usually designed with the small dimension less than $\lambda/2$ in order to pass waves of only one polarization—those with the electric field parallel to the short side. The phase velocity then depends on the other dimension.[1] This suggests that a wave guide should exhibit double refraction if both dimensions are greater than the cut-off value, $\lambda/2$. Waves polarized in the plane of the short side would progress down the guide at a lower velocity than those polarized parallel to the long side.

If a microwave transmitter and receiver[2] are available, an instructive demonstration piece may be easily constructed from a piece of sheet metal bent to form an almost-square wave guide about $0.6\times 0.6\times 10\lambda$. The joining edges of metal should be left free so that one dimension can be changed slightly by squeezing while in use. The guide is held in front of the transmitter with the guide axis parallel to the direction of propagation and a diagonal of the almost-square section parallel to the polarization plane of the incident radiation. When the guide is squeezed to minimize the signal received at the other end by an antenna parallel to the transmitting antenna, the receiving antenna may be rotated (by an assistant) to demonstrate the existence of a strong wave polarized at right angles to the incident wave.

The explanation follows precisely the theoretical discussion of an optical half-wave plate. The incident wave is resolved by the guide into two waves of nearly equal amplitude, polarized at right angles to each other. Due to the slight difference in dimensions, the two waves move at different velocities with a gradually increasing difference in phase as they travel down the guide. When the guide dimensions are varied until this phase difference amounts to 180° as the waves leave the guide, the waves recombine in free space to form a wave polarized in a plane rotated 90° from the original plane. Different amounts of squeezing give elliptically or circularly polarized beams. In the latter case, the guide is analogous to a quarter-wave plate.

Efficiency of the guide may be improved by snipping along the edges for a distance of 0.6λ from each corner and bending the square flaps out through an angle of about 30° to form a sort of horn at each end of the guide. This serves to gather a larger amount of energy from the incident wave and also to improve the directional properties of the output end.

* Now with the Sandia Branch of the Los Alamos Scientific Laboratory, Albuquerque, New Mexico.

[1] H. H. Skilling, *Fundamentals of electric waves* (Wiley, 1942), pp. 166–168.

[2] C. L. Andrews, *Am. J. Physics* **14**, 379 (1946).

Force Derived from Momentum and from Kinetic Energy

R. J. Stephenson
College of Wooster, Wooster, Ohio

In a recent issue of this journal Professor Hull[1] has given an interesting discussion of the principles of momentum and work applied to problems of the type of determining the force necessary to pull a uniform flexible rope from a heaped-up coil along a frictionless floor at a constant speed. This problem was originally propounded by Professor Sleator[2] under the heading of 'A new solution to an old problem.' Now if the force on the rope is obtained by taking the rate of change of momentum one gets twice the value of that obtained by using the relationship between work done and kinetic energy added. Professor Sleator comments on this difference in the following manner: 'Any small particle of the rope is set into motion suddenly, and, since the particle moves after the impact with the constant velocity of the already moving mass of rope, the impact is inelastic. In such an impact some mechanical energy is lost as heat, and if, as in this case, the mass of the particle picked up is very small compared to the mass already in motion, the energy so lost is equal to the resulting kinetic energy of the particle. Hence, the kinetic energy of the moving mass of rope represents only half of the necessary work done.'

Students frequently find this latter statement difficult to understand and the simple analogous problem given below has been found to clear up the difficulties. Suppose a block having a large mass M is moving with constant speed along a horizontal frictionless table on which are placed a number of equally spaced similar bodies each of mass m. As the large block moves along it makes inelastic impacts with the small masses so that a force F is required to maintain the constant speed. The large mass is similar to the portion of the rope already moving while the small masses represent the portion of the rope being set in motion.

At each impact between the mass M moving with constant speed v and one of the stationary small masses m there is an amount of momentum added equal to mv. If there are n masses in unit distance then there are nv impacts per unit of time and the time between two successive impacts is $1/nv$. Thus the force F required to keep the velocity constant is given by the momentum change per unit of time or $F=nmv^2$.

When each mass m is set in motion it receives an amount of kinetic energy $mv^2/2$, so that the kinetic energy added per unit of time is $nmv^3/2$. If this kinetic energy is added by a force F acting over the distance traveled per unit time v then $F=nmv^2/2$, or half the value given by the momentum equation. However, this has omitted the energy lost at each of the inelastic impacts. To calculate this we see that if no force were acting the velocity v_1 of the two masses after impact would be given by the principle of conservation of momentum as

$$Mv=(M+m)v_1, \text{ or } v_1=(Mv/M+m).$$

At each impact there would be a loss of energy of

$$\tfrac{1}{2}Mv^2-\tfrac{1}{2}(M+m)v_1^2=[Mv^2m/2(M+m)].$$

Since there are nv impacts per unit of time the loss in energy per unit of time is $Mmnv^3/2(M+m)$. Now if M is very much larger than m then this expression becomes $mnv^3/2$ which is the same as that received by the masses when set in motion. The loss in kinetic energy goes into heat at the inelastic impacts and work must be done to supply this energy. Thus the total work done by the force per unit of time is mnv^2, which gives the same value for the force as that obtained from momentum considerations. While this may appear to be an academic problem nevertheless it applies in such practical cases as that of finding the additional force and horsepower required by an engine when it is traveling at constant speed and picking up water from a trough between the rails, or the force required to set a freight train in motion when the cars are initially crowded together and, due to the looseness of the couplings, only one car at a time moves forward. This problem has been recently discussed by Professor Dadourian[3] who has analysed it in a somewhat different manner from that given above.

[1] G. F. Hull, *Am. J. Physics* **16**, 447 (1948).
[2] W. W. Sleator, *Am. J. Physics* **15**, 474 (1947).
[3] H. M. Dadourian, *Am. J. Physics* **16**, 344 (1948).

In fact, it would seem that only those who had some first-hand experience in the acquisition of new knowledge in some disciplined field would be able truly to appreciate how great the science of the past has been, and would be able to measure those giant accomplishments against their own efforts to penetrate a few millimeters further into the darkness that surrounds them.—J. R. Oppenheimer.

LETTERS TO THE EDITOR

Maxwell's Thermodynamic Relations

I SHOULD like to offer the following mnemonic for Maxwell's thermodynamic relations in reply to Professor Focken's[1] request:

Place the letters p, S, T, V in a square array in alphabetical order, i.e., in the order in which ordinary printed matter is read, thus

$$\begin{pmatrix} P & S \\ T & V \end{pmatrix}.$$

Next affix partial differential symbols, divide vertical elements, insert an equality sign, and affix the bottom element not used in the derivative for the subscript. This yields a first Maxwell relation,

$$(\partial P/\partial T)_V = (\partial S/\partial V)_T.$$

The other Maxwell relations are obtainable by interchanging diagonally opposite elements of the square array, each interchange introducing a negative sign.

Example: interchange T and S (introducing one negative sign), this leads to a second Maxwell relation as before,

$$\begin{pmatrix} P & -T \\ S & V \end{pmatrix}, \quad \left(\frac{\partial P}{\partial S}\right)_V = -\left(\frac{\partial T}{\partial V}\right)_S.$$

A third relation requires the interchange, $P \leftrightarrow V$ in the original array, and the fourth relation requires the double interchange $\begin{Bmatrix} T \leftrightarrow S \\ P \leftrightarrow V \end{Bmatrix}$ yielding a resultant positive sign.

Sharples Corp., Research Labs., 201 Spring Garden Street, Philadelphia, Pennsylvania

ROBERT E. PAYNE

[1] Charles M. Focken, *Am. J. Physics* **16**, 450 (1948).

Maxwell's Thermodynamic Relations

THE mnemonic proposed by Mr. Robert E. Payne is basically the same as the one described in my initial letter,[1] but in my opinion its form is preferable because two of the necessary steps, namely, the original array and the introduction of the negative signs, are easier to remember.

By combining the attractive features of various suggestions that have reached me, especially one from Mr. Herbert Rubin (College of Engineering, University of California), I have arrived at the following simple and satisfactory mnemonic for Maxwell's thermodynamic relations.

First write the square array of the letters P, S, T, V in alphabetical order, as proposed by Mr. Payne, thus

$$\begin{vmatrix} P & S \\ T & V \end{vmatrix}.$$

This may be considered as a square, which can be used erect, inverted, or turned on either side. These four positions yield the four equations when the square is divided into vertical elements, the partial differential symbols are affixed, the equality sign and the subscripts are added in the manner proposed by Mr. Payne. The negative signs are included when the square is used on either side.

University of Otago, Dunedin, New Zealand

CHARLES M. FOCKEN

[1] C. M. Focken, *Am. J. Physics* **16**, 450 (1948).

Measurement of Linear Coefficients of Expansion

IN a recent article by C. Williamson and W. H. Michener[1] an experiment is described whereby measurements of temperature *versus* expansion may be made on a brass tube using a dial gage as an indicating instrument. With several modifications this experimental arrangement has been utilized in the Massachusetts Institute of Technology freshman physics laboratory to measure the average linear coefficient of thermal expansion over the interval between the temperature of steam and that of cold tap water. The main advantage of the apparatus is the speed with which measurements can be made. If the tube is surrounded with some thermal insulating material, such as rock wool, a single determination of the coefficient of expansion can be made in less than ten minutes. Thus, a student is able to make measurements on several different tubes during the hour and a half allowed for the experiment.

The student first measures the coefficients of expansion of a brass tube and an aluminum tube. He is then given a unit containing a tube composed of a brass section and an aluminum section joined together. The total length of the compound tube is given but the individual lengths of the component parts are unknown. By measuring the linear coefficient of expansion of the compound tube and utilizing the previously measured coefficients of aluminum and brass, the student can solve two equations simultaneously to find the lengths of brass and aluminum in the compound tube. If the coefficients are measured to an accuracy of one percent, the length determinations can be made with an accuracy of 5 percent or better.

Finally, the student measures the coefficient of expansion of a glass tube blown from pyrex. The relatively small coefficient he finds is in agreement with his knowledge of the use of Pyrex cooking utensils.

Massachusetts Institute of Technology, Cambridge, Massachusetts

FRANCIS W. SEARS
NISSON A. FINKELSTEIN

[1] Williamson and Michener, "Some improved experiments for the heat laboratory," *Am. J. Physics* **16**, 160–161 (1948).

ANNOUNCEMENTS AND NEWS

Book Reviews

Cybernetics. NORBERT WIENER. Pp. 194, Figs. 8, 6¼ x 9¼ in., The Technology Press, John Wiley and Sons, Inc., New York, 1948. Price $3.00.

The twentieth century has often been called the electrical age. It is becoming increasingly clear that a more appropriate designation would be the age of communication and control. The phenomenal progress in telephonics and radio has been evident to all. Not so thoroughly realized by the general public is the rapid development of all sorts of devices, mainly electronic, for the precise control of vast amounts of energy in all branches of industry and in fact in large sectors of ordinary human activity. Few persons who drive a car ever stop to think of fundamental philosophical significance of the control gadgets which have been devised to make driving so easy. Even less thought has been given by most of us to the important fact that all such control mechanisms depend for their success on the sending of messages from one part of the system under control to another, *i.e.*, themselves exemplify the art of communication.

Though certain types of control mechanisms have long been utilized in engineering, it is only relatively recently that they have been subjected to intensive analysis under the general designation of *servomechanisms*. Examination discloses a fundamental similarity between the behavior of the ordinary ball governor on a steam engine, the common household thermostat, and the mechanical steering mechanism on a modern ocean liner. In each case a certain pattern of behavior (*i.e.*, speed of the engine, temperature of a room, direction of the boat) is maintained with a minimum of fluctuation by feeding back more or less continuously to the source of the behavior information as to how much the actual output is deviating from the desired pattern, and by providing means wherewith this information will lead to action in the correct (stabilizing) direction.

The use of the grid in a vacuum tube as a feedback mechanism for controlling the plate current is now thoroughly established in electronics. Through the work of the author of the volume under review and several of his colleagues, attention is now drawn to the fact that the nervous system of the living organism involves a feedback mechanism of essentially the same kind for the control of bodily movements. Thus it no longer seems likely that the carrying out of a purposeful act like the picking up of a book on the table is a simple one-way process in which the appropriate part of the brain dictates to the appropriate muscles by means of the neurons what action they must take in order to bring about the desired end. Rather every such purposeful act involves a circular process in which at each stage information is fed back to the central mervous system stating to what extent the originally desired end has failed to be achieved. The importance of this feed-back process becomes painfully evident in cases where it breaks down due to pathological conditions in the nervous system, *i.e.*, in various types of ataxia.

If the concepts of communication and control are thus so significant throughout the whole range of human activity, the mental discipline that studies them deserves a distinctive name. To coin this, Professor Wiener has gone back to the Greek for *steersman* (*cf.*, the significance of the *kappa* in Phi Beta Kappa), and the result is "Cybernetics." His book is a stimulating survey of the basic ideas of the new point of view with numerous specific applications.

Some of the applications will be fairly readily understood by the general reader. Thus he should have no serious trouble grasping the fundamental idea of feedback, though some of the accompanying mathematical analysis will baffle him. Moreover, the analogy between numerical computing machines on the binary scale and the human nervous system is clearly presented and provides a glimpse of the unifying tendency of cybernetics. On the other hand, the discussion of time series phenomena involved in the transmission of information against a background of noise makes considerable demands on the mathematical competence of the reader. The author introduces the definition of amount of information as the *negative* logarithm (to the base 2) of the ratio of *a posteriori* knowledge of the system in question to *a priori* knowledge. This has a close relation with entropy as defined in statistical mechanics: in fact it is the negative of the entropy, so that any increase in information we may gain about a system represents a decrease in entropy and vice versa. This association of gain in entropy, that is, increase in randomness, with loss in information is a highly illuminating one. There is an interesting application to the theory of Brownian motion which will repay some study by the physicist.

The book will doubtless make its greatest appeal through the discussion of the final chapters in which the author indulges among other things in speculations on the scanning techniques of the brain by which we recognize certain portions of experience as "wholes" or universals rather than as mere collections of parts: for instance, how we distinguish a human face viewed from any angle from that of a cat or dog. This leads to some interesting suggestions about sensory prosthesis or the replacement of the loss of one sense by the extra development of another. Psychopathology comes in for some attention and the analogy between the brain and a modern calculating machine is stressed. Both instruments sometimes break down and for possibly similar reasons; we should be careful not to overload either.

The importance of the transmission of information in maintaining equilibrium in any large system of separate organisms leads at once to a consideration of human society. The obvious failure of large social groups to attain equilibrium deprives the author of hope that the ideas of cybernetics will have much application to social problems. He indulges in the now well-nigh universal lament that technological progress has far outstripped social adjustment

and we may as well expect the worst. Such pessimism somewhat belies the buoyant optimism of the rest of the work with respect to the possibilities of the central idea.

This is a challenging book. Its principal theses will need further elaboration in somewhat simpler and more systematic form in order to drive home to the general public their possible significance for the culture of our time.

R. B. LINDSAY
Brown University

Science at War. J. G. CROWTHER AND R. WHIDDINGTON. Pp. 185+ii, Figs. 51, photographs 44 pages, 14½×22½ cm. Philosophical Library, Inc., New York, 1948. Price $6.00.

As Sir Henry Dale says in the foreword, this book was prepared under the sponsorship of the Scientific Advisory Committee to the British Cabinet, and is the first authoritative account of the scientific contribution to the war based on the official archives and documents of the Committee. The authors, Mr. J. G. Crowther, Chairman of the Association of British Science Writers, and Professor R. Whiddington, Head of the Department of Physics, University of Leeds, have written in a popular style using no mathematics whatever. The popularized technical explanations, however, are quite satisfying, and the technically minded reader will be glad to see the many quantitative statements. For instance, in regard to navigation by radar: "Aircraft controlled by Oboe and bombing from 30,000 feet at 250 miles drop on the average half of their bombs within 150 yards of the target."

The book is divided into four sections: Radar, Operational Research, The Atomic Bomb, and Science and the Sea.

For American readers the short section on the atomic bomb will be the least interesting, and may fall rather flat. Although the British had decided in the summer of 1941 that the construction of atomic bombs was feasible, and the early history of their project is interesting enough, this section of the book is concluded at the time (in 1943) when the British scientists left for the "great American Tube Alloys Establishment at Los Alamos." Evidently the use of meaningless titles for important projects was not confined to the USA.

The development of operational research is new. In a division on the optimum size of convoys, it is pointed out that "if escort vessels are to be out at a distance of 4000 yards from the convoy, the perimeter for a convoy of 80 ships would be one-seventh longer than that of a convoy of 40 ships. Thus seven escort ships can protect a convoy of 80 as well as six escorts can protect a convoy of 40." Elsewhere it is shown how the flying output of the Coastal Command was approximately doubled by changing the criteria for maintenance efficiency. There are interesting discussions of such questions as proper division of weight in an airplane between ammunition and radar gear, the placing of an additional gun on an aircraft at the expense of five miles per hour in speed, the relative advantages of spending effort in increasing performance of equipment or in training operators, etc.

The section on underwater science discusses magnetic and acoustic mines, and among other things describes the notable success of the anticipatory counter-measures work on pressure-change mines in relation to the Normandy invasion.

In the opinion of the reviewer, the section on radar, comprising about half the book, is really excellent. This stimulating story begins with the first radio-location experiments of the Heaviside layer in 1924, and continues with the following of aircraft by radio reflections in 1935, the construction of the Chain Home warning stations (1938–1941), the development of 10-centimeter radar (from 1939 onward), "Gee" (radio navigation based on time intervals between pulses from fixed ground stations, 1940–1942). "Oboe" (radar navigation based on radar control of groups of aircraft from two fixed radar stations, 1942), and "H_2S" (high power magnetron-equipped aircraft radar target-finders, 1942–1943). Passing mention is made of beacons, transponders, and an American radar gun-layer which combined with the British-invented-American-developed VT fuze was mainly responsible for the fact that "at the end of the attack, nearly 100 percent of all flying bombs approaching London were being shot down." It is significant that before the use of radio and radar navigation, during the first two years of the war "90 percent of the bombs dropped [by the British] had probably exploded harmlessly in the fields and had had no effect," and that the use in aircraft of not more than 50 sets of "H_2S" was primarily responsible for the victory over the submarine—monthly shipping losses decreasing from 700,000 tons in March 1943 to 100,000 tons in August.

There are more than a hundred charts, diagrams, and photographs, all of top quality, showing radar towers and antennas, magnetrons, radar and cathode-ray screens and pictures, damage from bombing raids, flame throwers, underwater explosions, cavitation behind ships' propellors, various mines, and so on. The numerous photographs add a great deal to the value of the book for the popular reader.

Perhaps three or four pages (no more) are concerned with nontechnical matters such as the democratic teamwork collaboration of the British scientific effort in contrast with the autocratic organization in Germany.

The reviewer feels that this book is not in the category of "must" reading, although like other enjoyable books he hopes it will be widely read because it is easy to read, authoritative, pleasantly informative, interesting, and occasionally exciting. At times it is grim. If further evidence is needed of how Britain narrowly escaped defeat by the timely development of new scientific weapons, this book gives the facts and figures of those developments.

There is one platitudinous paragraph to the effect that the "use of science as an aid to war is a perversion from its proper purposes, and its rapidly extending misuse . . . creates a threat to the survival of civilization." The reviewer does not condemn this paragraph for being a bromide (which it is), but he believes that any writer gifted enough to write a book on the subject of science and war should bear a responsibility to try to deal with some of the less concrete aspects of that subject such as the problems of the scientist (or of the nontechnical citizen) in his attempts to prevent war and promote peace. Some-

body *must* bring the discussion of these subjects into the open if effective answers are to be found. There are numerous moral, ethical, economic, political, and scientific arguments for actively attempting to prevent war. What better place could the scientist find for these arguments than in juxtaposition with his description of the tremendous technical accomplishments during the recent war both for preventing destruction and for creating it!

SEVILLE CHAPMAN
Cornell Aeronautical Laboratory

A Source Book in Greek Science. Edited by MORRIS R. COHEN AND I. E. DRABKIN. Pp. 579+xxi, Figs. 138, 16×23¾ cm. McGraw-Hill Book Company, Inc., New York, 1948. Price $9.00.

It has been said that the vigor of a state can be measured by the readiness with which it forgets its great men. If this proposition holds for fields of learning as well as for states, one might conclude that the sciences are passing into the softness of middle age, for a tide of interest in bygone scientists has been rising steadily for several decades. Such a gloomy conclusion can be side-stepped, however, by noting that the interest in the history of science is felt mainly by persons who themselves are not scientists. This fact suggests that science, by its impact on modern society, has finally forced itself to the attention of those who study human thought in the large. The scientist may take proper satisfaction in this belated recognition that human affairs have not been shaped by politics and literature alone. If he has a conscience, however, the teacher of science cannot relax in smug enjoyment of his enhanced importance. While his consequence has been heightened, his responsibilities have been broadened. Understanding the present state of his science is no longer enough: he must now be able to take part in the examination of science as an influence on society. In this examination the leaders will no doubt be specialists trained primarily in social studies and secondarily in science. The contribution made by the scientist will be twofold. He will use the general courses in science as vehicles for diffusing an understanding of scientific method, and he will make frequent allusion to the social and intellectual effects of various scientific discoveries. In short, he will treat science as a human activity. The end in view demands that the teacher of general courses in science become conversant with the history of his subject.

In many of the new courses designed for the general education of nonscientists, historical considerations are not neglected. One may ask whether the elementary and intermediate courses for prospective scientists cannot profitably include historical accounts of the great discoveries and conceptual innovations. The training of generals has long included a study of past military operations. Schlieffen's study of Hannibal's methods prepared the way for Ludendorff's victory at Tannenberg. If historical examples provide good training in the least sentimental of the professions, the managing of men in battle, can one dismiss them as valueless in the training of scientists—even applied scientists?

The scientist who spent much time studying his subject's past would rapidly become ignorant of its present. He needs a quick and painless way to survey the high spots of the past, and to settle specific questions as they arise in his teaching. Fortunately, there exist various compendiums of excerpts from books and papers by the great scientists. Outstanding in this class are the *Source Books in the History of the Sciences*, sponsored by the Carnegie Foundation and various learned societies, and published at intervals over the last twenty years under the general editorship of Gregory D. Walcott. The *Source Book in Physics* has doubtless saved the readers of this journal many a trip to the library. The volume under review is the most recent member of this series. In conjunction with some brief text like Singer's *A Short History of Science* (Oxford, 1941), it will enable the reader to acquire very quickly a useful degree of orientation concerning the scientific attainments of the Greeks. Now that such information is so easy to obtain, perhaps the absurd tradition that the Greeks had no real science will at last be snuffed out. Actually the work of Copernicus and Kepler, and some of the work of Galileo, rested directly on the astronomical and mathematical work of the Alexandrians. After a lapse of over a thousand years, the moderns continued from where the ancients left off.

The present volume has been edited by two men notable for the breadth of their learning. The late Professor Cohen, who like Whitehead was trained as a mathematician before turning to philosophy, wrote extensively on logic and scientific method as well as on the philosophy of the law. Dr. Drabkin combines mathematics and a knowledge of Greek and Latin with some years of postdoctoral study in the history of science; many of the translations are his own.

The book covers a wide range of sciences, the round number of pages devoted to each being as follows: Mathematics 90, Astronomy and Mathematical Geography 90, Physics 170, Chemistry 20, Geology and Meteorology 20, Biology 70, Medicine 60, Psychology 30. At the end is a bibliography useful for more specialized study. Most of the excerpts are short, which makes the book ideal for evening browsing. It is also suitable for more sustained reading, but its usefulness for reference will be lessened by the want of an index of subjects. The extensive index of names will perhaps suffice for the expert, but will be of limited value to one not already acquainted with the field. On the other hand, the copious and well-conceived notes render great service to the neophyte. Even within a given section, the extracts are not arranged chronologically, but a chronological perspective can be established with the help of the name-index and the biographical notes listed therein. Some of the translations are from Latin authors; the title of the book is justified by the fact that such science as the Romans had was taken from the Greeks.

In the section on physics, the excerpts are taken mostly from the works of Aristotle, Euclid, Archimedes, Ptolemy, Lucretius, Hero, Vitruvius, and Simplicius. Aristotle's work is useful chiefly as an example of how not to develop physics. Later, when the experimental approach was better established, real progress was made in statics and optics,

although there were, of course, many errors. In the final period the work ran mostly to gadgetry, much of it devoted to priestly and military purposes. Perhaps the most surprising excerpts—representing the swan song of Greek physics—are Philoponus' flirtation with the concept of inertia, and his clear description of the falling-body experiment for which Galileo won fame 1100 years later.

J. J. G. McCue
Smith College

Vacuum Manipulation of Volatile Compounds. R. T. Sanderson, Pp. 162+viii, 40 Figs, 5½×8½ in. John Wiley and Sons, Inc., New York, 1948. Price $3.00.

The author's preface to this little book states that, since literature on the methods and apparatus involved in high vacuum work are not readily obtainable in integrated form, "it has been difficult for one to begin to use this technique without learning it from another individual. It is the purpose of this book, therefore, to provide in one place enough practical information to enable an investigator with an average background of scientific training and experience to construct and operate a general purpose high vacuum apparatus for chemical research." Let it be at once said that the book fulfills its stated purpose admirably; its usefulness to other people and other groups will be discussed below.

The author develops his subject in 15 chapters covering 115 pages. To this are added several appendices, totaling 35 pages, and an author and subject index. The chapters are Introduction to Chemical High Vacuum Technique (2 pages); Materials of Construction and Operation (11 pages); Glass Blowing (16 pages); Producing a High Vacuum (13 pages); Measuring Pressure (15 pages); Low Temperatures (14 pages); Valves (4 pages); Introducing Into and Removing from Vacuum System (6 pages); Measuring Volume (10 pages); Measuring Vapor Pressure (4 pages); Condensation Traps and Fractionation (8 pages); Miscellaneous Suggestions and Methods (10 pages); Practice in High Vacuum Technique (6 pages); A General Utility High Vacuum Apparatus (5 pages); and Future Developments (2 pages). The appendices include data for vapor pressure thermometers filled with carbon disulfide, sulfur dioxide, ammonia, carbon dioxide, hydrogen chloride, ethylene, methane, oxygen, and nitrogen; a reference table for copper-constantan thermocouples; and the vapor pressures of almost 400 pure compounds at the melting point of carbon disulfide, the sublimation point of carbon dioxide, the atmospheric boiling point of ammonia, the melting point of water, and at 20.0°C.

The development of most of the subjects presented is so brief as to preclude mention of many alternative techniques, let alone a discussion of certain more modern methods. The author, apparently trained in the school of Stock, Schlesinger, Burg, and co-workers, leans heavily on their techniques. These, of course are admirable for the development in the fields in which they worked, and can also be used in most instances in general high vacuum manipulation. However, modern research has led to the development of new and important methods, which are omitted or only briefly mentioned in the book. Thus, for instance, there is no discussion of vacuum gauges other than the classical McLeod type. The use of oil diffusion pumps is only indicated in passing, and no mention is made of the new silicone oils that have been developed as pumping fluids. Metal and glass-metal techniques are entirely omitted. Throughout the book the emphasis rests on manipulations involving the minimum contact of chemicals in vacuum systems with ground joint and stopcock lubricants and sealing materials other than mercury. While this stress has been necessary in most of the work emanating from the Stock group, it would appear that in many instances the judicious use of these materials leads to easier and more rapid manipulations than can be obtained with the awkward mercury seal techniques described. The utility of ground joints, both conical and spherical, cannot be overestimated in the reviewer's mind, but these have been given only passing consideration in Sanderson's book.

Some rather qualitative statements are made, which cannot help but leave the impression of lack of scientific rigor. Thus density is given without units; it is stated that there are "innumerable collisions" of gas molecules at N.T.P.; the section on principles on condensation suffers from lack of statement of assumptions, leading to questionable statements. However, the general clarity of presentation of the manipulative techniques overbalances these objections, leaving an over-all favorable impression of the book.

Two sections of the book require high commendation. The reviewer, having seen illness due to mercury poisoning and death due to failure to observe precautions in handling liquid air, can only urge most strongly that all workers in the field read the corresponding sections and follow to the letter the instructions given in them. If for no other reason, Sanderson's book should be generally valuable.

While the extended table of vapor pressures of pure compounds in the appendix will be valuable to a certain very specialized group, the general utility of data on such esoteric compounds as carbon sulfoselenide and 1,3-dichlorotetramethyldisiloxane is questionable. Here again the background of the author is evident. Undoubtedly, experimenters getting so deeply into a given field will look beyond the appendix into the literature for further data. On the other hand, the data on vapor pressure thermometers and copper-constantan thermocouples is quite generally useful.

It is the reviewer's feeling that students and research workers doing their first jobs in vacuum manipulation will derive real benefit from the use of this book. More experienced groups will gain as always from refreshing their memories, but will not find many new "wrinkles" to try.

J. D. Heldman
Wilmington, California

Fear, War, and the Bomb. P. M. S. Blackett. Pp. 238+vii, 14½×21½ cm. Whittlesey House, New York, 1949. Price $3.50. (Published in England under the title *Military and Political Consequences of Atomic Energy.*)

Professor Blackett has probably written the most important book that has yet appeared on the political and

military aspects of atomic energy. What he says may not be correct, but it appears that his views are essentially the same as those of the political leaders of the U.S.S.R.; in presenting his case Professor Blackett therefore gives the reader much understanding of what has occurred in the United Nations discussions of atomic energy control during the past three years. This is not to imply that the book is a superficial propaganda tract. On the contrary, Professor Blackett writes on a basis of historical documentation, and with complete realism, in a manner that is consistent with his having achieved what he has in the realm of physics.

There are three major themes in the book. First, that long range *strategic* bombing (in contrast to *tactical* bombing which is integrated with land and sea attacks) is not of primary military importance, even though atomic bombs are used. Second, that the purpose of the American use of atomic bombs at Hiroshima and Nagasaki was not primarily to bring to a close the second world war, but rather to achieve an initial political victory in the third world war. And third, that the Russian rejection of the American proposals for international control of atomic energy is altogether justified by Russia's legitimate military and economic self-interests. It seems to me that the degrees of validity which Professor Blackett establishes for the three themes are in an order of rank about the same as the above order of listing.

The discussion of the military significance of strategic bombing is based largely on the effects of Allied bombing of Germany and Japan in the recent war, as reported by the United States Strategic Bombing Survey. Such facts as that German war production rose from a base line of 100 in 1940 to 285 in 1944, in spite of "saturation raids" on the war production centers, are convincing evidence for the thesis that the use of atomic bombs will not in itself provide an easy way to defeat a great power. The sober, factual discussion which Professor Blackett gives of the role of atomic bombs in achieving military objectives is perhaps the most valuable aspect of this book. The reader will not be led to feel that atomic bombs are any net gain to the world's well-being, but he will see the error of those who see immediate destruction of our great cities as immediately consequent to the start of a war with a Russia that possesses atomic weapons.

The semiofficial American justification for the atomic bombing of Japan has been that the great losses of American and Japanese lives which invasion would have entailed were thereby avoided. Professor Blackett lays considerable stress on the advice given to the War Department by the Manahattan Project scientists committee, of which James Franck was chairman, to the effect that atomic bombs should not be dropped on Japan. The Franck report was re-enforced by a petition to the same effect, signed by sixty-four scientists of the Chicago Metallurgical Laboratory. Why was this advice not heeded? Professor Blackett can find no reason among the given military reasons. He points out that Japan was all but prostrate: America had air control, even over the home islands; and the U.S.S.R., in accordance with agreements made at Yalta, was preparing to attack the large Japanese land forces in Manchuria. American invasion was not scheduled to begin until November 1st, and hence, Blackett argues, there was hardly reason to drop atom bombs on Japanese cities early in August; for there was likelihood that by November the Japanese would have surrendered even without invasion. Actually, the bombs were dropped on August 6th and 9th, and apparently they precipitated the Japanese surrender of August 14th. The Russian offensive in Manchuria began on August 9th. Blackett's assertion is that the atomic bombs were dropped so that Japan would surrender to the United States before Russia had become extensively involved in operations against the Japanese. The occupation of Japan thereby became an exclusively American activity. By the annihilation of a large part of the population of two cities, the assertion concludes, the United States cleverly achieved a great diplomatic and political victory.

Professor Blackett's defense of the Russian attitude toward the Lilienthal-Baruch proposals for an international Atomic Development Authority seems to rest on two chief points: (1) that such an authority would probably prevent Russia, power-poor country, from developing atomic power plants as rapidly as it could if unrestrained by the authority; (2) that Russia would be in a relatively weak position militarily during the transition period to complete international control; for example, the location of military and industrial plants would of necessity become known to international inspectors. (There would be no like disadvantage for America, Professor Blackett soberly tells us (p. 153), because at present there is little secrecy anyway about the location of essential plants in America.) The reader will find the arguments for the above points to be presented in detail and with great clarity by Professor Blackett. Not all readers will be persuaded to sympathy for the Russian attitude, but one can scarcely fail to gain understanding of that attitude.

Many things could be said in rejoinder to various details of Professor Blackett's *apologia pro Gromyko*. I will only, however, comment on the attitude that man is incorrigibly nationalistic, which the author seems to bring to his subject. This attitude is well illustrated by his comment (p. 158) that the Baruch plan ". . . must be considered historically as an astute move and a very considerable victory for American diplomacy," this on account of the great "success with which this specious plan was put across in most countries as a wise and generous measure." Professor Blackett's world is one in which political leaders think altogether in terms of the advantage of their own country and distrust of the motives of other nations. There is, it appears, no force at all behind international good will, except when it serves power politics to use the sentiments of well meaning but innocent idealists.

Professor Blackett may be correct in his attitude. If so, his book is doubly useful. If he is wrong, it is still useful, for he has clearly apprised us of the nationalistic and power politics thinking that in any case is a major force in political activity today. In wondering whether or not it is the only force, I am sadly reminded of the recent remark of Dr. Albert Einstein on the subject of world peace: "Where there is no will there is no way."

RICHARD SCHLEGEL
Michigan State College

Proceedings of the American Association of Physics Teachers

The New York Meeting, January 27–29, 1949

THE eighteenth annual meeting of the American Association of Physics Teachers was held at Columbia University on January 27–29, 1949. PROFESSOR H. K. SCHILLING was chairman of the program committee for the meeting. Local arrangements were in charge of PROFESSOR J. R. DUNNING.

A banquet with the American Physical Society was held on Thursday evening, January 27, in the Grand Ballroom of the Hotel New Yorker. Speakers at the dinner were PROFESSOR ROBERT E. CUSHMAN on the subject, "Freedom versus Security in the Modern World," and PROFESSOR P. W. BRIDGMAN on the subject, "Fundamental Democracy and the Forgotten Physicist."

Invited Papers and Reports

Papers and Special Events

Techniques for demonstrating color phenomena. ISAY A. BALINKIN, *University of Cincinnati.*

Presentation of the Oersted Medal of the American Association of Physics Teachers to Professor Arnold Sommerfeld. PAUL KIRKPATRICK, *Chairman of the Committee on Awards,* and J. W. BUCHTA, *President of the Association.*

Acceptance of the Oersted Medal on behalf of Professor Sommerfeld. E. U. CONDON, *National Bureau of Standards.*

How World War II has affected the science of physics—seventh Richtmyer Memorial Lecture of the American Association of Physics Teachers. LEE A. DUBRIDGE, *California Institute of Technology.*

Joint Session with the American Physical Society

Fields and quanta—address of the retiring president, American Physical Society. J. R. OPPENHEIMER, *Institute for Advanced Study.*

Panel Discussion: Physics in General Education

Physics in general education—keynote address. JAMES B. CONANT, *Harvard University.*

Aims and schemes in general education in physics courses. ERIC ROGERS, *Princeton University.*

Physics courses in the general education program at Harvard. P. LE CORBEILLER, *Harvard University.*

Avoidable dangers accompanying the rapid development of general education. G. W. STEWART, *University of Iowa.*

Physics in physics. A. G. SHENSTONE, *Princeton University.*

Panel Discussion: On the Training of College Teachers of Physics

The need for better training of college teachers of physics. T. R. MCCONNELL, *University of Minnesota.*

How should graduate schools modify the training patterns for prospective college teachers? G. B. PEGRAM, *Columbia University.*

Should prospective college physics teachers be required to do research? K. LARK-HOROVITZ, *Purdue University.*

Training college physics teachers to aid in improving science teaching in secondary schools. P. G. JOHNSON, *U. S. Office of Education.*

Symposium: Testing in Physics Teaching

Testing in terms of specific objectives of physics teaching. LEO NEDELSKY, *University of Chicago.*

Tests being developed at Colgate. C. L. HENSHAW, *Colgate University.*

Testing for critical thinking in physics. PAUL J. BURKE, *Educational Testing Service.*

Proposal for Cooperative Testing Program

C. N. WALL, on behalf of the *Committee on Testing* of the American Association of Physics Teachers.

Contributed Papers, with Abstracts

Two sessions were devoted to the following contributed papers:

1. A borderline problem between physics and biology: selfduplication of genes. HERBERT JEHLE, *University of Pennsylvania.*—The most striking example of specific attraction between identical large molecules is the selfduplication process of genes, *i.e.*, of chromosomes in mitosis. From the surrounding medium, the constituent molecules of the original gene collect molecules which happen to be identical with the former ones, placing them near to the corresponding molecules of the original gene. Thereafter chemical bonds link these molecules together, forming a daughter gene. What is so astonishing in this phenomenon is its accuracy of differentiation which permits identical replicas to be handed down through enormous numbers of chromosome generations; its versatility which implies that all kinds of genes and their mutants selfduplicate accurately; and the long range of the specific forces involved (the interacting identical macromolecules are usually separated by other constituent molecules of the gene.) Thus we have to look at this phenomenon as something very different from ordinary chemical bond formation. Classical anharmonic resonance between thermally excited identical macromolecules throws some light on this problem.

2. Friction, a brief review, two new theorems and a problem. G. P. BREWINGTON, *Lawrence Institute of Technology.*—Amonton's Law ($f=\mu N$) quantitatively describes friction phenomena. The source of (dry) friction has been explained by two classes of theories as due (a) to rubbing of irregular surfaces and (b) the "stick slip" process. It has also been observed that between two well-cleaned or outgassed surfaces the coefficient of friction may become as high as 4.00. It is proposed that between two large moving surfaces in contact, the energy dissipated is divided between the two materials proportionally to their heat conductivities. Also, that for clean surfaces the coefficient of friction is greater for materials having the same crystal lattice constant. An interesting problem occurs when one flywheel, initially at rest, is brought up to speed by engaging by means of a friction clutch, with another initially rotating flywheel. The heat lost is independent of the coefficient of friction in the clutch.

3. The bicycle ergrometer. LESTER I. BOCKSTAHLER, *Northwestern University.*—The electrodynamic bicycle ergrometer often used in physiology laboratories lends itself very well for experimental study in the physics laboratory. It should be especially helpful in teaching premedical students. Such an ergrometer will be described. Construction plans, constants, methods for calibration, and use will be given in detail. The elementary principles of electricity as well as the characteristics of the direct current motor and generator are, of necessity, brought out clearly in the use of this machine. It also serves to tie up and coordinate many principles and units of mechanics and electricity. Cooperative projects between the departments of physics and zoology may be carried out with mutual profit.

4. Simple demonstration experiments. G. M. KOEHL, *George Washington University.*—Simple demonstration experiments are very often more valuable as teaching aids than the more involved and spectacular ones in which the principle being demonstrated is often confused because of the students' interest in the auxiliary equipment. If the demonstrations are simple, many students will want to repeat them at home as "After Dinner Science." This stimulates further interest in science, and is also an excellent learning process, for the student must have a clear understanding of the principles involved in order to explain the phenomena to his "audience". Several demonstrations will illustrate this. First the images formed by plane mirrors will be demonstrated and studied. Are these images "reversed left to right" as so many textbooks state? Next the classical question, "What happens to the water level in a lake when a metal ship sinks in that lake?" will be answered by a demonstration. The magnitude of the atmospheric pressure is usually convincingly demonstrated by means of Magdeburg hemispheres. How can it be demonstrated without the use of the hemispheres, pumps, or other laboratory equipment? Surface tension can be demonstrated with nonlaboratory equipment as can many of the phenomena of heat.

5. A method of teaching alternating current circuits. J. GIBSON WINANS, MALCOLM COLE, GERALDINE ROSS WALTERS, VIRGINIA HAZELWOOD HUMMEL, *University of Wisconsin.*—The method of simplification of circuit works for both a.c. and d.c. circuits if all additions are vector instead of scalar. Kirchoff's laws for a.c. circuits are: (1) The vector sum of currents approaching any point in a network equals the vector sum of currents leaving that point. (2) In any closed circuit the vector sum of electromotive forces equals the vector sum of the potential drops. The directions for current and voltage are horizontal right for pure resistive impedance, up for pure inductive impedance, and down for pure capacitive impedance. From Kirchoff's laws one derives $Z=Z_1+Z_2+Z_3\ldots$ for impedances in series and $1/Z=1/Z_1+1/Z_2\ldots$ for impedances in parallel. Impedance Z and $1/Z$ are taken as vectors in the same directions as voltages and currents. Impedances are really scalars, but the resultant impedance may be determined by vector addition. With these equations most a.c. circuits simplify to a series circuit and then to a single electromotive force E and single external impedance Z. The current $I=E/Z$ is the current for the series circuit. With this current all potential drops may be determined. With all potential drops known each current in the original circuit can be determined by $I=V/Z$. The currents for several impedances in parallel must add vectorially to equal the series circuit current.

6. Combining of simple electronic instruments into a Z meter and its use in studying characteristics of radio equipment. O. L. RAILSBACK, *Eastern Illinois State College.*—The individual instruments combined are standard decade resistance, oscillator electronic switch, and oscilloscope. The impedance to be measured is placed in series with the standard resistance and the output of the signal generator applied across both. The respective voltages across standard and unknown are applied to the inputs "A" and "B" of the electronic switch and hence to the oscilloscope. The setting on the standard resistance is adjusted until both signals stand at the same height on the scope. The reading on the standard is then the measured impedance. The phase angle can also be read on the scope which permits the computation of the reactive and resistive components as well as the determination of Q. The apparatus has been found useful for observing phase relations and measuring impedance of loud speakers at various frequencies. Impedance *vs.* frequency curves for speakers are shown. It is also useful in demonstrating the resistive character of a mechanical load on the speaker diaphragm, the phase angle shifting from nearly zero to about sixty degrees in some cases by merely stopping the speaker diaphragm with the hand. The "reflection" of load from secondary to primary in a transformer can be shown.

7. Demonstration of Pohl's interference experiment. MARK W. ZEMANSKY, *City College of New York.*—The two virtual images of a source due to reflections from the two surfaces of a thin mica plate may be used as coherent sources for the production of circular interference fringes, as first shown by R. W. Pohl and described on p. 68 of

Pohl's *Einfuhrung in die Optik*. A small high pressure mercury arc is used as a source, and a three in. square of mica 0.003 in. thick as a reflector. Interference fringes covering a whole wall of a lecture hall may be produced in this way and demonstrated to a large audience.

8. The Fundamentals of Colorimetry. DONALD J. LOVELL, *Norwich University*.—The chromaticity diagram is derived and the dominant wavelength, purity, and luminous reflectance of a substance are found on this diagram. These qualities are the physical attributes of color and compare to the psychological factors of hue, saturation, and brightness. Color measurements are applied to many industries, and basic knowledge of modern color measurement procedures should be had by all physics teachers. The psychological attributes of color are explained by a hypothetical observer who sorts various colored samples. The samples are distributed by hue, saturation, and brightness. Hue determines the characteristic of a color by which we call it red or blue, etc. Saturation determines the ratio of mixture between a "pure" color and a gray. Brightness is found by comparing the sample to a gray scale. The discussion also includes the development of the tristimulus values and the Young-Helmholtz theory of color vision. From a luminous reflectance curve of a sample, the luminousity curve of the human eye, and the spectral distribution of the light source, chromaticity coordinates are derived. These are entered into a chromaticity diagram such as the one found in the *Handbook of Colorimetry*[1] and the physical factors found. An example is followed through.

[1] A. C. Hardy, *The handbook of colorimetry* (Technology Press).

9. A laboratory experiment on trajectories. FLOYD W. PARKER, *Lincoln Memorial University*.—It is well known that a sphere rolling on an inclined plane surface moves in a parabolic path. An elaboration of this basic idea has resulted in a laboratory experiment in which projectiles are simulated by rolling steel spheres which move in parabolic paths on the surface of a slightly tilted laboratory table. The dilution of g associated with the tilting of the table enables an observer to follow the motion of the projectile with ease. A simple launcher will be exhibited which is used to give the sphere a known and repeatable initial velocity v_0. Preliminary experiments involve the determination of v_0 and of the acceleration of the pseudo-gravity effective in accelerating the sphere. With these values known, the values of the time-of-flight T, range R, and maximum height attained H, are computed for various angles of elevation. Verification of the formulas used and of the accuracy of the computations is accomplished by means of trial shots with small targets placed at the computed positions of H and R. The value of T is verified with a stopwatch. Suggestions pertaining to such items as laboratory procedure, surface smoothness, frictional effects, and target size will be made, and information relative to the accuracy obtainable, based on the percentage of hits achieved by approximately 300 laboratory students, will be presented. A short auxiliary experiment will also be described.

10. Observations of a "reactionary" physics teacher. HENRY A. PERKINS, *Trinity College*.—Inability of students to get information from the printed word and inadequate training in mathematics are our most serious obstacles which can only be overcome by better training in the schools. However, in spite of this unhappy state of things, much can still be accomplished at the college level. The vital discipline of deriving an important formula, now too much neglected, can still be required. It is better to memorize a proof, if necessary, than never to read it at all. A certain amount of oral recitation is advocated, with reservations. The inductive method in elementary laboratory classes is discouraged. The chief value of such a course is acquiring technique and method rather than theory, and there should be a careful explanation of the experiment before the student performs it. Finally, types of examination are discussed, and opinions given on how to grade them, closing with a defense of reactionary policies in general when we need to recover certain values that have been lost through progress in a wrong direction.

11. The contribution of physics laboratory work to general education. G. E. OWEN, *Antioch College*.—While the physics laboratory should contribute to the general education of the student by making him familiar with some phenomena, some instruments, and some techniques which will contribute to his effectiveness as an individual, its major contribution has to do with developing some skill in applying the scientific method and a disposition to use it. The ordinary form of laboratory work based on technical experiments, modified for the general student, gives him no experience in some of the most important aspects of the scientific method. He does not formulate the question, plan the experiment, or even arrive at an unbiased conclusion (he arrives at the conclusion he is led to expect is the "right" one). He does get a chance to interpret the data, but in many cases the way in which he is to make the interpretation is prescribed and the end result must agree with a theory or a handbook. It is important that the student develop a disposition to apply the scientific method, so far as it can be applied to all his problems. The physics laboratory work should be so designed that he gets a chance to use it, and successfully, or that disposition will not be encouraged. Several experiments are described which will enable the student to gain some of the values described.

12. A special science laboratory for nonscience education. ERNA M. J. HERREY, *Queens College*.—Since more adequate understanding of science is now generally considered indispensable for the layman-citizen, nobody doubts that nonscience students should receive instruction in science, though opinions differ on the *how*. But, whether students should take a conventional laboratory course or special *science surveys*, or courses based on historical or philosophical approach, the following are always major difficulties which the nonscience student must be helped to overcome: The intimidating immensity of the subject; students' prejudices regarding specific abilities required to grasp science; lack of familiarity with scientific matters

and even elementary mathematics; habits of trying to learn by reading and memorizing instead of understanding; prevalence of interest in passing marks rather than in the subject itself. The essential point seems: to make the student learn through his own immediate experience with tasks of science. Lectures and book reading should be reduced in favor of simple experiments selected to familiarize with purpose and methods of science rather than detailed factual results. This was tried out by introduction of laboratory periods specially devised for nonscience students. Independently, each student had to solve simple problems through self-planned experimentation with simple apparatus and unbiased interpretation of observed results. Examples of suitable experiments are given and this method of achieving understanding of science compared with other methods.

13. College physical science courses in general education. C. C. CLARK, *New York University.*—Assuming that the idea that physics (or physical science) courses for general education should have a place in college curricula is generally accepted, the following questions present themselves. What are the primary objectives of such courses? What shall their content be? What methods of teaching shall be used? These courses should give the student some insight into the scientific method as a way of handling problems, and provide him with some understanding of physical science phenomena. The subject matter to be studied is to be carefully selected to include some of the fundamental principles of the physical world, particularly those whose applications affect so closely and extensively our daily experiences. It should include a few well chosen historical examples of the development of science, presented so as to illustrate the rational and impartial methods used in solving scientific and related problems. All subject material is to be on a nontechnical level, devoid of extensive and higher mathematics. It should be fully illustrated in the classroom with appropriate and working demonstrations. A course based upon the above ideas has been offered by New York University for twenty years. Taken by several thousand students, it has met with satisfactory results.

14. A course for nonscience majors. W. J. JACKSON AND E. A. TOWNSEND, *Rutgers University.*—The course sets before the student ceratin broad objectives, namely, to gain some *understanding* of the physical world as made known by science; to become familiar with the concepts and habits of thought which have produced modern physical science and the manner in which these concepts and habits have been used to solve certain problems; and finally to come to some understanding of the significance of science and its application in our modern society. Recognizing that such broad objectives might lead to one of two extremes, either a very descriptive course *about* science or the typical general course with only minor changes, carefully considered procedures have been used in developing this course. Briefly, these are: (*i*) a careful selection of subject matter, (*ii*) an introduction of certain material from the fields of astronomy and chemistry to give a basis for the development of certain ideas later, (*iii*) the introduction of certain significant ideas in their original form through materials now becoming more readily available to enable the student to get an historical perspective, (*iv*) a cutting across the traditional classification of fields whenever necessary, through classroom work and collateral reading, to have the student grasp the significance of the growth of concepts through man's experimentation and development of theory, (*v*) a selection of laboratory experiments carefully designed to help with the understanding of the important concepts studied and to give added emphasis to the importance of measurements in the growth of physical science, (*vi*) the introduction of more advanced group experiments, too difficult for the general student to do alone, but necessary in attaining the objectives. The content of the course and the teaching devices used will be discussed in some detail.

15. A refresher program for high school physics teachers. ELMER HUTCHISSON, *Case Institute of Technology.*—As a contribution toward the better training of scientists in this country, the General Electric Company has, during recent summers, established fellowships for high school science teachers at Union College and at Case Institute of Technology. At Case, fifty fellowships, all for teachers of physics, were granted in each of the summers of 1947 and 1948. These fellowships included tuition, board and lodging, and travel expenses to and from Cleveland. The program of study, under the direction of R. S. SHANKLAND, R. C. PUTNAM, L. O. OLSEN, and E. C. GREGG, was designed to provide a review of the fundamental concepts of physics and of recent developments in the physical sciences, so that the fellows could return to their classrooms at the end of the summer aware of the importance of strong emphasis on fundamentals and of the directions in which physical science is progressing. It was expected that their teaching should reflect this renewed knowledge, and they should be more likely to awaken scientific ambition in the minds of students.

16. Heat and thermodynamics in elementary courses. ERIC M. ROGERS, *Princeton University.*—While mechanics is taught in a systematic way to provide a needed background, heat is usually taught with a disappointing choice of topics and emphasis. The treatment seems to be a survival from some more extensive course. A plea is made for reconsideration of the syllabus, to give less attention to some topics, such as thermal expansion, and more to others, such as kinetic theory and perhaps thermodynamics. If thermodynamics is treated at all it should be carried to a stage where the student can see its use—for example, the Clausius-Clapeyron relation should be obtained. A device for discussing Carnot efficiency is shown; and a scheme is given for developing the Kelvin scale easily.

17. Did the Greeks perform experiments? OTTO BLÜH, *University of British Columbia.*—The discussion of the historical development of physical science in college courses for nonscientists is now accepted as an important element

of the so-called 'cultural approach'. In future greater recognition will also be paid to the history of physics in the curriculum of the science major, and especially of the physicist. It will then be significant to deal with the subject not in a superficial and sweeping manner, and without any of the popular preconceptions. Contrary to the aknowledged speculative achievements of the Greeks, their scientific work, until recently, has been treated in a disparaging manner, and modern scientists still adhere readily to the view that Greek physical science did insufficiently appreciate quantitative observation and experimentation, which are claimed to be modern practices. The question is supposed to be obscured by the lack of information, but classical literature leaves us not without indication that planful observation and experiment were at the basis of Greek scientific statements. The notion that the Greeks were averse to experimenting because this involves manual labor reflects more the attitude of the 18th and 19th century gentleman-philosopher than the original Greek way of thinking. They could have satisfied their admittedly high curiosity for knowledge by employing their slaves in the execution of experiments. The spherical shape of the earth, the models of planetary motion, the laws of acoustics, or such considerations as were included in their atomistic theory of matter, are still frequently presented as speculative achievements, based on 'aesthetical arguments', 'number speculation', and the play of pure imagination. We can hardly believe that such knowledge is found so easily, that such truths are discovered by inspired vision rather than by accurate observation, collection of data, planful experiments, and theoretical generalization; and there is no cogent necessity for this belief on account of sufficient evidence to the contrary. But the question remains why we have no literary sources which give direct information about Greek experimental work and working methods. The explanation is to be found, in my opinion, in their attitude to scientific work in general. Not only manual work, but also the work of the intellectual, the artist, poet, and scientist, was considered a calamity, and despised by the noble Greek. Nevertheless, it was not considered undignified for the artist and scientist to do his work, because he was considered to submit himself to an overpowering urge originating in superpersonal or divine powers. Intellectual work therefore had to have the stamp of inspiration and vision. And when scientific results were obtained the hard way, the methods had to be obscured, and the pretension upheld that the results had been obtained through inspiration. Such conventions made it undesirable to describe toilsome observational and experimental work. The Greek scientist's attitude to 'publications' was therefore fundamentally different from ours, which, at times, is more concerned about the exposition of scientific methods than about the results of research. It is not astonishing, for the mentioned reason, that written evidence of experimental work has not been preserved. In trying to follow the intricacies of classical thought in connection with considerations on the history of science, wider educational values can be evolved, serious misunderstandings dispelled, and obsolete statements eventually avoided.

18. Reactance measurements with a "Lecher wire" system—an experiment in advanced electrical measurements. T. B. BROWN, *George Washington University.*—The reactance to be measured is made the termination of a resonant line (parallel wires) and its value then computed from the characteristic impedance of the line and the distance from the termination to the nearest voltage-node. A pair of wires several meters long is stretched between fixed points and divided into two sections by connections to a short lateral line which couples it to the oscillator. One section serves as the tuned line which is terminated by the reactor while the other section is provided with a shorting bridge and serves as a tuning stub. The connections to the lateral line make sliding contact with the main line so that the oscillator may be placed anywhere along its length; this permits easy operation of the line over a wide frequency band. Readings may be made with high precision, and systematic errors have been reduced to small values by careful design. This experiment has been very useful in the advanced undergraduate laboratory where it serves to correlate ordinary resonance methods of measurement at RF with VHF methods employing concentric lines and wave-guides. It has much of the simplicity of the former while it involves all the basic principles of the latter method.

19. WHEN is a physicist? STANLEY S. BALLARD, *Tufts College.*—In view of the much-heralded emergence of physics as a profession, regarding which all of us are enthusiastic, serious thought must be given to setting up professional standards applicable to individuals. The physics professor must be provided with an answer for his undergraduate majors when they ask, "When can I say that I am a professional physicist?" At present there is in this country no ready-made answer to this question in the field of physics, as there is in the various fields of engineering and perhaps also in chemistry. In view of the rapidly increasing place that physics is establishing for itself in industrial research, development, and production control, the Ph.D. degree can no longer be accepted as the necessary criterion for having attained professional stature in physics, at least in other than academic circles. Presumably some acceptable combination or combinations of formal educational training and subsequent professional or *on the job* experience must be worked out. This may involve the accrediting or certifying of college and university physics departments and/or their curricula, a procedure which will be fraught with many practical difficulties. The results of an informal questionnaire taken among Tufts College physics majors will be presented as an illustration of current student attitudes as to when a person should be able to call himself a professional physicist, and some of the many possible combinations of educational training and professional experience which might suffice to certify a professional physicist will be discussed briefly.

Report of the Secretary

The Executive Committee held its annual meeting in the American Institute of Physics Building, January 28, 1949. President Buchta presided.

The following persons were present: *Mildred Allen, Henry Barton, C. E. Bennett, *D. M. Bennett, *O. H. Blackwood, *L. Bockstahler, *J. W. Buchta, *B. H. Dickinson, H. L. Dodge, J. D. Elder, *R. C. Gibbs, E. Hutchisson, *C. Hodges, H. Hughes, E. C. Kemble, *Paul Kirkpatrick, *G. M. Koehl, K. Lark-Horovitz, *R. W. Lefler, Harvey Lemon, *W. V. Norris, *C. J. Overbeck, *R. R. Palmer, *R. F. Paton, *W. B. Pietenpol, *O. L. Railsback, *E. M. Rogers, *Duane Roller, *H. K. Schilling, F. G. Slack, R. M. Sutton, M. H. Trytten, B. B. Watson, M. W. White, and A. G. Worthing. Asterisks indicate the names of voting members at this meeting. The others were present by invitation.

Business with American Institute of Physics.—Our dues for 1949 to the American Institute of Physics, as a supporting member society, have again been set at 15 percent of the 1948 dues income. Our increased membership gives us four representatives on the Governing Board of the Institute. The following are nominated to represent A.A.P.T. for the terms stated: Marsh White 1947–50; P. E. Klopsteg 1948–51; Paul Kirkpatrick 1949–52; H. K. Schilling 1949–52.

The Journal.—Upon recommendation of Editor T. H. Osgood, the following appointments have been made. Assistant Editor—B. H. Dickinson; Associate Editors (three year term)—J. E. Goldman; H. N. Otis; Eric Rodgers.

Reports of officers and committees.—The work of the officers is demonstrated through the meetings and general activity of the Association. A study of the Treasurer's report, published elsewhere in this issue, shows that we operate on a firmer financial footing than ever before. A number of new projects are, therefore, being initiated now to give the membership the benefit of this resource.

The membership committee under Dr. Marsh White has been primarily responsible for an increase of some seven hundred new members during the past year, bringing the present total to over 2650 members. Time and the increased membership have focused attention on a need for revision of portions of the Constitution and By-Laws. Three committees have been working on phases of this job. It is probable that their final report will be presented to the members of A.A.P.T. for action within this year.

We continue our cooperation through representatives with other organizations (A.C.E., A.A.A.S., N.S.T.A., A.S.E.E., etc.). Some of these representatives are on important joint committees. All of our own committees are active. The results of the work of our committee on Letter Symbols and Abbreviations will be found in several science hand books. The "Coulomb's Law Committee" has just completed a hundred page report which will soon appear in print. The committee on Visual Aid to Teaching is working on four projects which will be of service to our membership. The first of these is expected to take definite form during 1949.

The L. W. Taylor Memorial Fund.—The untimely death of L. W. Taylor is a severe loss to the Association. An anonymous friend of Dr. Taylor has started a Memorial Fund in which others may wish to participate.

Future meetings.—The American Society for Engineering Education again invites the A.A.P.T. to cooperate in holding a June meeting this year at Troy, New York, June 21–25. The Executive Committee has accepted this invitation and also plans to hold a meeting June 27–29 at Seattle, Washington. The details for the next annual meeting have not yet been worked out.

Election of officers.—The count of the 777 ballots cast in the December election has resulted in the election of the following new officers:

President: (1 year term) J. W. BUCHTA.
Vice-President: (1 year term) DUANE ROLLER.
Secretary: (2 year term) R. F. PATON.
Treasurer: (2 year term) P. E. KLOPSTEG.
Executive Committee: (2 year term) R. B. BRODE, F. G. SLACK.

Attendance at annual meeting.—The A.A.P.T. attendance for this past meeting was most encouraging. Nearly every session filled the rooms to capacity and on some occasions even standing space was not available after over 300 auditors had filled both seats and aisles. The prepared speeches and the floor discussions which followed were equally stimulating.

Annual business meeting.—The eighteenth annual business meeting was held Saturday, January 29th at 11:15 A.M. A brief report of A.A.P.T. activities was given by the Secretary, followed by reports from committee chairmen There was no business requiring official action of the members present.

A vote of thanks was extended to the Secretary now completing the Constitutional limit of six years service and to the Program chairman, Vice-President H. K. Schilling, for his excellent work in arranging a very successful program.

C. J. OVERBECK
Secretary, 1943–1949

Annual Report of the Treasurer

Balance brought forward from December 15, 1947 (See auditor's report for 1947)		$ 5,128.74
CASH RECEIVED		
Dues received for 1948	$11,963.10	
Dues received for 1947	5.00	
U. S. Treasury Bond Interest	200.00	
Royalties from McGraw-Hill	262.50	
Receipts for AAPT dinner reservations—1948 meeting	693.00	
L. W. Taylor Fund	250.00	
Total deposited, 12–14–47 to 12–31–48		13,373.60
Total cash available		$18,502.34

DISBURSEMENTS

Payments to American Institute of Physics		
Net balance for AIP publication*	3,030.31	
Other charges in connection with publication	2,366.75	
15% of dues to support of AIP	1,240.10	
Purchase of Treasury Bonds, Sept. 8, 1948	5,085.55	
January 1948 meeting, payment to caterer	690.00	
Postage and supplies, Editor's Office	438.05	
Travel expenses, Editor and Assistant Editor	241.03	
Assistants, Editor's Office	1,567.00	
Expenses of Secretary's and Treasurer's Offices	569.02	
Cooperative Committee	59.03	
Expenses of President's Office	42.06	
Expenses of Membership Committee	1,832.57	
Expenses of Nominating Committee	87.35	
American Council on Education Dues	100.00	
American Council on Education, travel expenses of AAPT representative	102.60	
Miscellaneous	49.20	
Total disbursed		17,500.62
Balance on hand December 31, 1948		$ 1,001.72
The Association held at December 31, 1948, U. S. Government Treasury Bonds and Notes having a par value of		$15,000.00

* Includes $1,385.45, payment for operations between September 30 and December 31, 1947.

I have audited the books of account and records of Dr. Paul E. Klopsteg, Treasurer of the American Association of Physics Teachers, for the year ended December 31, 1948, and hereby certify that the foregoing statement of receipts and disbursements correctly reflects the information contained in the books of account. Receipts during the year were satisfactorily reconciled with deposits as shown on the bank statements, and all disbursements have been satisfactorily supported by vouchers or other documentary evidence. U. S. Government Bonds and Notes of $15,000 par value are held in safekeeping by the State Bank and Trust Company of Evanston, Illinois, and a certificate was obtained from the custodian as of December 31, 1948.

WILLIAM J. LUBY
Certified Public Accountant

Chicago, Illinois

In connection with the formal report of the Treasurer, which is herewith submitted, the following comments are of interest:

Because of the large number of nonmember subscriptions, the amounts paid towards publication costs by authors, and the substantial sale of back numbers, the *American Journal of Physics* had a net income of $9,822 from sources other than membership dues. Its total cost to the Association of $13,713 was thus reduced to $3,891. This total cost is just a few cents under $6 per year on the present membership basis so that each member benefits to the extent of about $4.20 by virtue of the other sources of income mentioned.

It has been the policy of the Treasurer through the years to try to anticipate financial needs on a long-term basis. Much experience with organizations that are constantly struggling with financial problems made him aware of the desirability of provision for advancing costs, an expanded program of publication, and for such nonpublication activities as might be desirable for the Association to undertake. One example of this is the financial requirement of an active committee such as the Membership Committee, which in 1948 disbursed $1,833. (That its efforts were highly successful is shown by the net increase of 666 members in good standing at the close of the year.) Provision also had to be made to relieve the Secretary's and Treasurer's offices of some of the burdensome detail that goes with rising membership. During the years past practically all of this work had been carried by these two offices without expense to the Association. A substantial part of that expense continues to be carried by the institutions with which these officers are connected. During 1948 the collection of dues, maintenance of the mailing list, and the addressing and mailing of notices, as well as other miscellaneous routine, were handled by the American Institute of Physics at a cost of $2,740. This includes publication of the membership directory and reprinting of back issues for which a continuing demand exists.

The Executive Committee in its actions has wisely implemented the planning of the Finance Committee. Through the years the Association has been successful both in accomplishment of its aims and in its finances. It has accumulated a reserve of about $15,000 which constitutes a substantial guarantee towards the undertaking of any worthwhile projects that may seem desirable. Beginning with 1948 our dues were increased from $5 to $6. Notwithstanding that this was only 20 percent—a very much lower increase than has been necessary in many other societies—that the number of issues of the Journal was increased from 6 to 9, and that there was unusual expense in the transfer of the editorial office, we succeeded in breaking even for the year. Out of the $1000 excess of receipts over disbursements, about $500 is still to be paid the American Institute of Physics for operations in the fourth quarter of 1948. Thus there is an estimated net surplus from the year's operations of about $500.

The Association, in common with the other member societies, is fortunate in its close relationship with the American Institute of Physics. This connection has provided the members with a subscription to the new journal, *Physics Today*, without additional cost, although there may be the necessity of making a small charge in the future. It also provides the excellent services that the Association could otherwise secure only with difficulty and probably at substantially greater cost.

It is gratifying to the Treasurer to be able to present a report as optimistic as this one. It has been especially gratifying to see the steady and substantial growth of the Association since its inception. It has been responsible for creating and developing through its very competent editors a journal which has filled a serious void among the

publications in Physics, and which is highly appreciated by many subscribers who are not members of the teaching profession. In the society the teachers of physics have found a meeting ground where they can, without hesitation or apology, discuss problems important to their profession. The strength of the society, in numbers, activities, and finances, assures an even more successful future.

PAUL E. KLOPSTEG, *Treasurer*

DIGEST OF PERIODICAL LITERATURE

1 Rutherford = 10^6 Disintegrations per Second

The determination of the number of radioactive atoms in a sample of a radioactive isotope can only be accomplished by using procedures that will yield absolute measurements, or by using uniform standards of comparison that will give the same result whenever the determination is made. In the measurement two factors are involved: One is that the methods which are correct in principle be used; the other is that a reasonable degree of accuracy in producing and using the standards must be attained.

What is desired in the measurement of a sample of a radioisotope is the number of radioactive atoms present. This determines the strength of the source. The actual number of radioactive atoms present N is related to the rate at which atoms are disintegrating by the disintegration constant λ, according to the fundamental law of radioactive decay

$$dN/dt = -\lambda N.$$

It is, therefore, necessary to measure the *rate* of disintegration in order to determine the *number* of atoms present in a sample. This disintegration rate is generally equal to the number of nuclear particles, usually beta-particles or positrons, which are emitted per unit time from all atoms disintegrating within this time interval.

Most of the difficulties in making quantitative determinations of radioisotopes can be eliminated, if standard sources of the radioisotope under measurement are available. Unfortunately, it is possible to produce standards of only a relatively few radioisotopes. Isotopes having large disintegration constants and known disintegration schemes can, however, be compared with other isotopes with fair accuracy if (1) it is known that a beta ray or a positron is emitted for each disintegration, and (2) the maximum energy of the beta-ray spectrum of the standard is not too different from that of the isotope to be measured.

For many purposes a unit for the disintegration rate which is smaller than the curie and which can be specified independently of the decay rate of radium is desirable. The name *rutherford* (abbreviation rd) has been suggested by E. U. Condon and L. F. Curtiss[1] for a rate of 10^6 disintegrations per second. The millirutherford and the microrutherford are convenient submultiples of the unit. While the disintegration rates of radioisotopes vary over a wide range, the unit is of convenient size. For example, therapeutic doses of many radioisotopes are of the order of 100 rd, while the weakest source that can be measured by the mica-window bell-type counters is about 1 μrd. Radioactive standards available from the National Bureau of Standards include Co^{60} solution in 5 ml ampoules rated at 1.5 rd and 0.15 rd. LEON F. CURTISS, "Measurement of Radioactive Isotopes," National Bureau of Standards Circular 473 (1948). (Circular may be obtained from the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C., at 5 cents per copy.)—B.H.D.

[1] E. U. Condon and L. F. Curtiss, *Physical Rev.* 69, 672 (1946).

New Members (*continued from page 221*)

Hammond, T. M., Schreiner Institute, Kerrville, Texas
Healy, John C. (J), 3702 Gulf Ave., Houston, Texas
Hoyt, Rosalie C., Dalton Hall, Bryn Mawr College, Bryn Mawr, Pa.
Irland, Max J., 916 N. Rosevere, Dearborn, Mich.
James, John, Sr., Box 227, Northport, Ala.
Johnson, Russell Hughes, 3802 Oak Ave., Baltimore 7, Md.
Keene, Ellis L., No. 5 East North St., Dover, Del.
Knapp, E. J., 2100 N. Stanton St., El Paso, Texas
Knox, Kenneth E., Parsons Junior College, Parsons, Kansas
Langholz, Samuel A., 1078 East 12th St., Brooklyn 30, N. Y.
Loferski, Joseph J. (J), Physics Dept., University of Pennsylvania, Phila., Pa.
Mange, Phillip Warren (J), 1216 Merrill St., Kalamazoo, Mich.
May, Albert, 3915 Livingston St., N.W., Washington 15, D. C.
Miller, Richard Edgar (J), 448 N. Austin Blvd., Oak Park, Ill.
Montgomery, Charles Edward (J), 577 Oxford St., London, Ontario, Canada
Morgan, John H. (J), 250 Cheapside St., London, Ontario, Canada
Muni, Harry Walter, Hartwick College, Oneonta, N. Y.
Nedelsky, Leo, Examiner's Office, University of Chicago, Chicago 37, Ill.
Odell, Gerald Paul, Compton College, Compton, Calif.
Orfuss, Irving, 444 East 20th St., Apt. 5B, New York 9, N. Y.
Owen, Wilfred Robert, 168 First St., Romeo, Mich.
Patten, Robert R. (J), H-1 Stearns Village, Medford, Mass.
Ramsey, Allan, 3401 Sophia Way, Sacramento 17, Calif.
Reid, Harry A., Alabama Agricultural & Mechanical College, Normal, Ala.
Rittenburg, Stanley Edward (J), 121 Glenville Ave., Allston 34, Mass.
Roane, Asa Em., (J), Rt. 1, Box 136B, Berrien Springs, Mich.
Roche, Thomas E., Jr. (J), 1013 Walnut St., Newton Highlands 61, Mass.
Scheffler, Edward C. (J), 56 Cornelia Ct., Roxbury 20, Mass.
Schleicher, Robert B. (J), 3477 Rosendahl Ave., San Diego 10, Calif.
Sladky, Richard E., 1402 W. Locust St., Milwaukee, Wis.
Smith, Stephen Judson (J), 1708 W. Michigan Ave., Kalamazoo, Mich.
Storer, James E., 376 Woodbridge Ave., Cambridge, Mass.
Tracy, Fred, 521 E. Beaver, State College, Pa.
Treanor, Charles E. (J), 51 Ketchum Place, Buffalo 13, N. Y.
Vuylsteke, Arthur A. (J), 1162 Lakewood Ave., Detroit, Mich.
Walton, Joe R., 123 Canton Hall, Oak Ridge, Tenn.
Whitney, James A. (J), 338 East 4th St., National City, Calif.
Wickenden, Roger C., R.R. No. 2, Oxford, Ohio
Wilkins, Lawrence E., 2738 Knox Ave., Los Angeles 26, Calif.
Williams, Perry W., Physics Dept., South Dakota State College, Brookings, S. D.
Woodhams, Raymond T. (J), 56 Forward Ave., London, Ontario, Canada
Wray, Shirl D., R.R. No. 7, Crawfordsville, Ind.
Zieman, Clayton M. (J), Wabash College, Crawfordsville, Ind.